

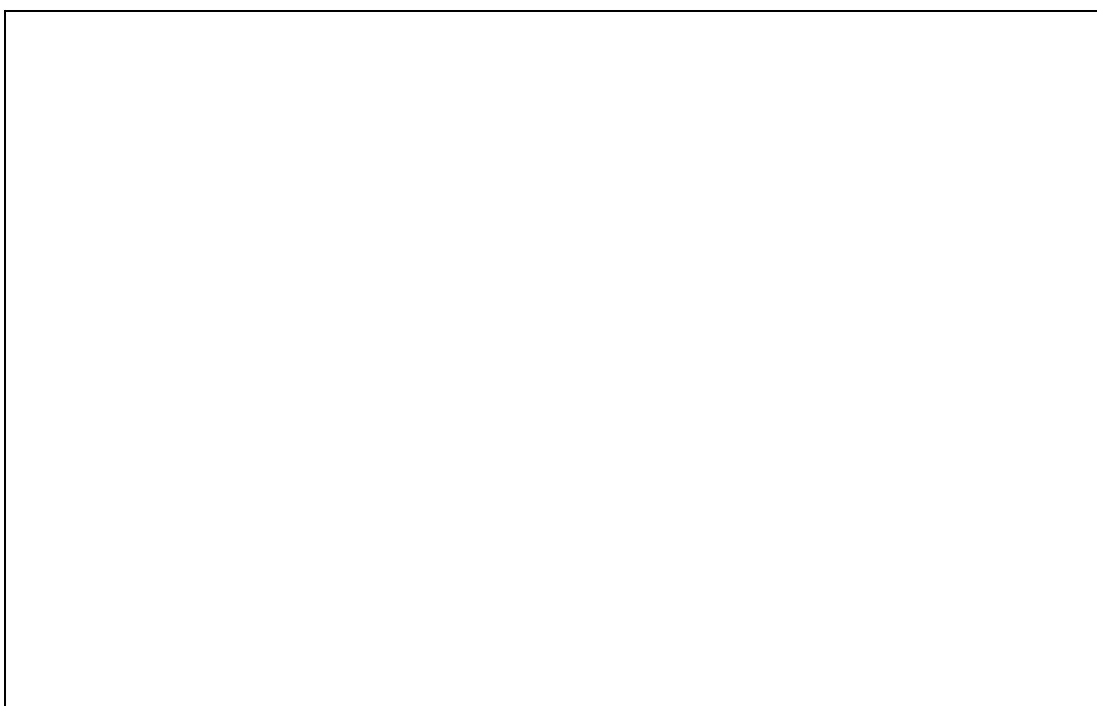
FÁBIO HENRIQUE JUGLER
FÁBIO ROBERTO STELLA SANTOS
JACKSON AUGUSTO PROSTT
RICARDO ASSUNÇÃO VIALLE

SISTEMA INTEGRADO PARA ANOTAÇÃO AUTOMÁTICA DE GENOMAS

Trabalho de Conclusão de Curso apresentado ao Curso de Tecnologia em Análise e Desenvolvimento de Sistemas da Universidade Federal do Paraná como requisito à obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Professor Dr. Roberto Tadeu Raittz

CURITIBA
2013



RESUMO

Apresentamos um sistema web para anotação de genomas de procariotos integrado a ferramenta SILA. O sistema tem por objetivo, fornecer o serviço da anotação automática de genomas em ambiente web, oferecendo opções para visualização e análise das informações. O sistema utiliza HTML5 para representação gráfica interativa das informações de anotação, fornecendo visualização do mapa genômico e de agrupamentos por produtos de proteínas, funções e por organismos vizinhos baseado nos genes anotados. Além das informações obtidas a partir anotação do SILA, o sistema associa dados do repositório de enzimas ENZYME. O sistema está disponível através do página do programa de pós-graduação em Bioinformática da UFPR pelo endereço www.bioinfo.ufpr.br/sila.

Palavras-chave: Bioinformática, Anotação automática de genomas.

ABSTRACT

We present a web system for annotation of prokaryotic genomes integrated with the tool SILA. The system aims to provide the service of automatic annotation of genomes in a web environment, offering options for visualization and analysis of information. The system uses HTML5 to interactive graphical representation of annotation information, providing visualization of genomic map and gene clusters by protein products, functions and neighbor organisms based on annotated genes. Besides the information obtained from SILA, the system combines the data repository of enzymes, ENZYME. The system is available through the website of the Bioinformatics course of UFPR by the url www.bioinfo.ufpr.br/sila.

Key-words: Bioinformatics, Automated genomic annotation.

LISTA DE FIGURAS

FIGURA 1 - FORMATO FASTA	17
FIGURA 2 - EXEMPLO DE REGISTRO NO FORMATO GENBANK	18
FIGURA 3 - EXEMPLO DE INFORMAÇÕES ASSOCIADAS NA ANOTAÇÃO	21
FIGURA 4 - TELA DE LOGIN	29
FIGURA 5 - TELA DE CADASTRO	30
FIGURA 6 - TELA PRINCIPAL	30
FIGURA 7 - TELA DE TAREFAS	31
FIGURA 8 - TELA DE DETALHES DA ANOTAÇÃO	32
FIGURA 9 - TELA DISTRIBUIÇÃO POR COG	33
FIGURA 10 - TELA DISTRIBUIÇÃO POR PRODUTO	33
FIGURA 11 - TELA DISTRIBUIÇÃO POR ORGANISMO	34
FIGURA 12 - TABELA DE GENES AGRUPADOS	34
FIGURA 13 - TELA DETALHES DO GENE	35
FIGURA 14 - TELA INFORMAÇÕES DA ENZIMA	35
FIGURA 15 - DIAGRAMA DE CASOS DE USO	40
FIGURA 16 - DIAGRAMA DE TRANSIÇÃO DE ESTADOS	41
FIGURA 17 - DIAGRAMA DE COMPONENTES	42
FIGURA 18 - DIAGRAMA DE ATIVIDADES	43
FIGURA 19 - DIAGRAMA DE CLASSES	44
FIGURA 20 - DIAGRAMA ENTIDADE RELACIONAMENTO	45
FIGURA 21 - DIAGRAMA DE SEQUÊNCIA - CADASTRAR USUÁRIO - FLUXO PRINCIPAL	69
FIGURA 22 - DIAGRAMA DE SEQUÊNCIA - CADASTRAR USUÁRIO - FLUXO A1	70
FIGURA 23 - DIAGRAMA DE SEQUÊNCIA - CADASTRAR USUÁRIO - FLUXO A2	71
FIGURA 24 - DIAGRAMA DE SEQUENCIA - REALIZAR LOGIN DE USUÁRIO - FLUXO PRINCIPAL	72
FIGURA 25 - DIAGRAMA DE SEQUENCIA - REALIZAR LOGIN DE USUÁRIO - FLUXO A1	73
FIGURA 26 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO PRINCIPAL E A6	74
FIGURA 27 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A1	75
FIGURA 28 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A2	76
FIGURA 29 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A3	77
FIGURA 30 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A4	78
FIGURA 31 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A5	79
FIGURA 32 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO PRINCIPAL	80
FIGURA 33 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A1	81
FIGURA 34 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A2	82
FIGURA 35 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A3	83
FIGURA 36 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A4	84
FIGURA 37 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A5	85
FIGURA 38 - DIAGRAMA DE SEQUENCIA - VISUALIZAR INFORMAÇÕES DA ANOTAÇÃO - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11 e A12	86
FIGURA 39 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTO POR COG - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10 e A11	87
FIGURA 40 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO PRINCIPAL	88
FIGURA 41 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A1	89
FIGURA 42 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A2	90
FIGURA 43 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A3	91
FIGURA 44 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A4	92
FIGURA 45 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A5	93
FIGURA 46 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR ORGANISMO - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10 e A11	94

FIGURA 47 - DIAGRAMA DE SEQUENCIA - VISUALIZAR INFORMAÇÃO DO GENE - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9 e A10 95

FIGURA 48 - DIAGRAMA DE SEQUENCIA - ALTERAR CONTA DO USUÁRIO - FLUXO PRINCIPAL, A1, A2, A3, A4 e A5 96

LISTA DE TABELAS

TABELA 1 - CODIFICAÇÃO DE AMINOÁCIDOS	15
TABELA 2 - CATEGORIAS FUNCIONAIS DO COG	19

LISTA DE ABREVIATURAS E SIGLAS

BAM	Formato de arquivo para representar alinhamento de sequencias
BLAST	Basic Local Alignment Search Tool
CDS	Coding DNA Sequence
COG	Clusters of Orthologous Groups
DNA	Ácido desoxirribonucleico
EC	
NUMBER	Enzyme Comission Number
EMBL	European Molecular Biology Laboratory
ExPASy	Expert Protein Analysis System
FASTA	Formato utilizado para armazenar sequências de bases e de aminoácidos em arquivo texto
GBK	Tipo de arquivo que armazena anotações
GenBank	Banco de dados público do National Center for Biological Information
HTML	HyperText Markup Language
IUBMB	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
JRE	Java Runtime Environment
JSP	JavaServer Pages
JVM	Máquina Virtual Java
MATLAB	Matrix Laboratory
mRNA	RNA Mensageiro
NCBI	National Center for Biotechnology Information
NR	Banco de proteínas não redundante disponibilizado pelo NCBI
ORF	Open Reading Frame
PDB	Protein Data Bank
PIR	Protein Information Resource
PRF	Protein Research Foundation
RGS	Rapid Gene Searcher
RNA	Ácido ribonucleico
rRNA	RNA Ribossômico
SIB	Swiss Institute of Bioinformatics
SILA	Ferramenta de anotação de genomas
tRNA	RNA Transportador
UML	Unified Modeling Language
W3C	World Wide Web Consortium
WBS	Work Breakdown Structure
XML	Extensible Markup Language

SUMÁRIO

1 INTRODUÇÃO	12
1.1 CONSIDERAÇÕES INICIAIS	12
1.2 OBJETIVOS	13
2 FUNDAMENTAÇÃO TEÓRICA	14
2.1 BIOINFORMÁTICA	14
2.2 CONCEITOS BIOLÓGICOS	14
2.2.1 Anotação	16
2.3 ARMAZENAMENTO E OBTENÇÃO DE INFORMAÇÕES BIOLÓGICAS	16
2.3.1 FASTA	16
2.3.2 GenBank	17
2.3.3 COG	19
2.3.4 Pfam	19
2.3.5 ExPASy ENZYME	20
2.4 SILA	20
2.5 JSP	21
2.6 HTML5	22
2.6.1 Scribl	22
2.6.2 RGraph	22
2.7 JQUERY	23
2.8 UML	23
2.9 MATLAB	23
3 METODOLOGIA	25
3.1 MODELO DE PROCESSO DE ENGENHARIA DE SOFTWARE	25
3.2 PLANO DE ATIVIDADES	27
3.3 PLANO DE RISCOS	27
3.4 RESPONSABILIDADES	27
3.5 MATERIAIS	27
3.6 DESENVOLVIMENTO DO PROJETO	27
4 APRESENTAÇÃO DO SOFTWARE	29
4.1 ACESSO AO SISTEMA	29
4.2 ENVIAR ARQUIVO PARA ANOTAÇÃO	30
4.3 TAREFAS SUBMETIDAS	31
4.4 DETALHES DA ANOTAÇÃO	31
4.5 AGRUPAMENTOS	32
4.6 DETALHES DOS GENES	34
5 CONSIDERAÇÕES FINAIS	36
REFERÊNCIAS	38
APÊNDICES	40
1 DIAGRAMA DE CASO DE USO	40
2 DIAGRAMA DE TRANSIÇÃO DE ESTADOS	41
3 DIAGRAMA DE COMPONENTES	42
4 DIAGRAMA DE ATIVIDADES	43
5 DIAGRAMA DE CLASSES	44
6 DIAGRAMA ENTIDADE RELACIONAMENTO	45
7 ESPECIFICAÇÃO DE CASOS DE USO	46
7.1 UC001 – CADASTRAR USUÁRIO	46
7.2 UC002 – REALIZAR LOGIN DE USUÁRIO	48

7.3 UC003 – ENVIAR NOVA TAREFA	50
7.4 UC004 – VISUALIZAR TAREFAS DO USUÁRIO.....	52
7.5 UC005 – VISUALIZAR INFORMAÇÃO DA ANOTAÇÃO	54
7.6 UC006 – VISUALIZAR AGRUPAMENTOS POR COG	57
7.7 UC007 – VISUALIZAR AGRUPAMENTOS POR PRODUTO.....	59
7.8 UC008 – VISUALIZAR AGRUPAMENTOS POR ORGANISMO	61
7.9 UC009 – VISUALIZAR INFORMAÇÃO DO GENE.....	63
7.10 UC010 – ALTERAR CONTA DO USUÁRIO	66
8 DIAGRAMAS DE SEQUÊNCIA	69
8.1 UC001 - CADASTRAR USUÁRIO	69
8.2 UC002 - REALIZAR LOGIN DE USUÁRIO.....	72
8.3 UC003 - ENVIAR NOVA TAREFA	74
8.4 UC004 - VISUALIZAR TAREFAS DO USUÁRIO.....	80
8.5 UC005 - VISUALIZAR INFORMAÇÕES DA ANOTAÇÃO	86
8.6 UC006 - VISUALIZAR AGRUPAMENTOS POR COG	87
8.7 UC007 - VISUALIZAR AGRUPAMENTOS POR PRODUTO	88
8.8 UC008 - VISUALIZAR AGRUPAMENTOS POR ORGANISMO	94
8.9 UC009 - VISUALIZAR INFORMAÇÃO DO GENE	95
8.10 UC010 - ALTERAR CONTA DO USUÁRIO	96
9 PGP – PLANO DE GERENCIAMENTO DE PROJETO	97
9.1 INTRODUÇÃO.....	97
9.1.1 ESCOPO E PROPÓSITO DO DOCUMENTO	97
9.1.2 OBJETIVOS DO PROJETO	97
9.1.2.1 OBJETIVOS.....	97
9.1.2.2 FUNÇÕES PRINCIPAIS	97
9.1.3 RESTRIÇÕES TÉCNICAS E ADMINISTRATIVAS.....	98
9.1.4 ESTIMATIVAS DO PROJETO.....	98
9.1.4.1 DADOS HISTÓRICOS USADOS NAS ESTIMATIVAS.....	98
9.1.4.2 TÉCNICAS DE ESTIMATIVAS	98
9.1.4.3 ESTIMATIVAS	99
9.1.5 RISCOS DO PROJETO.....	100
9.1.5.1 ANÁLISE DOS RISCOS	100
9.1.5.2 IDENTIFICAÇÃO	100
9.1.5.3 AVALIAÇÃO DOS RISCOS	101
9.1.5.4 ADMINISTRAÇÃO DOS RISCOS.....	103
9.1.5.5 OPÇÕES PARA EVITAR OS RISCOS	103
9.1.5.6 PROCEDIMENTOS DE MONITORAÇÃO DOS RISCOS.....	103
9.2 CRONOGRAMA.....	104
9.2.1 WORK BREAKDOWN STRUCTURE	104
9.2.2 GRÁFICO DE GANTT	105
9.3 RECURSOS DO PROJETO.....	106
9.3.1 PESSOAL	106
9.3.2 HARDWARE E SOFTWARE	106
9.4 ORGANIZAÇÃO DO PESSOAL.....	106
9.4.1 ESTRUTURA DA EQUIPE	106
9.5 PLANO DE COMUNICAÇÃO	107
9.6 MECANISMOS DE RASTREAMENTO E CONTROLE	107
9.6.1 VERIFICAÇÃO E CONTROLE DO ESCOPO.....	107
9.6.2 CRONOGRAMA	108
9.6.3 QUALIDADE	108
9.6.4 GERENCIAMENTO DA EQUIPE DO PROJETO	108

9.6.5 RELATÓRIO DE DESEMPENHO.....	108
9.6.6 GERENCIAMENTO DAS PARTES INTERESSADAS	109
9.6.7 RISCOS	109
9.6.8 MUDANÇA DE ESCOPO	109

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

A anotação de genomas é um processo complexo que envolve diferentes aspectos com o objetivo de obter informações estruturais e/ou funcionais sobre uma ou várias sequências relacionadas a um determinado genoma. No processo de anotação genômica, uma sequência desconhecida é documentada através da identificação de vários sítios e segmentos envolvidos na funcionalidade do genoma (ROUZÉ; PAVY; ROMBAUTS, 1999).

A anotação consiste em duas etapas principais: a predição de genes, onde realiza-se a identificação de elementos, tais como, ORFs (*Open Reading Frames*), estrutura gênica, regiões codificantes e localização de motivos reguladores; e a agregação de informação, que identifica funções bioquímicas, biológicas, regulações, interações e expressões. Ambas podem envolver experimentos biológicos e análises *in silico*. Existem diversas ferramentas para auxiliar o especialista humano a realizar essas tarefas, poupando-lhe tempo e esforço (STEIN, 2001).

Ferramentas de anotação automática visam diminuir o esforço gasto com tarefas rotineiras possibilitando a direta análise dos resultados. Dessa forma, sistemas desse tipo adquirem o caráter de serviço e precisam estar disponíveis aos pesquisadores a qualquer momento. Um sistema web proporciona a disponibilização de serviços de forma imediata, sem a necessidade de uma preparação a priori do ambiente de trabalho. Assim, o sistema aqui apresentado busca fornecer um serviço de anotação automática de genomas e disponibilizar possibilidades de análise e interpretação dos dados gerados, apresentando boa ergonomia, rapidez e facilidade de utilização.

Neste trabalho apresentamos e documentamos os elementos de desenvolvimento de um sistema web para anotação de genomas. O sistema proposto fornece visualização e análise de resultados em formas gráficas, controle de usuários e gerenciamento de tarefas. O sistema é integrado com a ferramenta de anotação automática de genomas chamada SILA (VIALLE, 2013, em fase de elaboração).

1.2 OBJETIVOS

O objetivo geral do projeto é fornecer um sistema em plataforma web integrado ao software de anotação automática de genomas SILA, oferecendo opções para visualização e análise das informações.

O sistema possui os seguintes objetivos específicos:

- Produzir uma interface de entrada de dados para o sistema de anotação;
- Interpretar os arquivos GBK gerados pelo sistema de anotação;
- Transferir as informações para um banco de dados relacional;
- Prover opções para gerenciamento de usuários;
- Fornecer interface para controle de tarefas dos usuários;
- Relacionar os dados das anotações com as informações do banco de enzimas;
- Desenvolver formas de agrupamento de genes baseado em informações da anotação;
- Prover diferentes opções de visualização das informações geradas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 BIOINFORMÁTICA

A Bioinformática é a ciência que utiliza abordagens computacionais para responder questões biológicas a partir de grandes e complexos conjuntos de dados (BAXEVANIS; OUELLETTE, 2001). Assim, através de programas de computadores, busca fazer inferências biológicas derivando predições úteis e interessantes (LESK, 2002).

Duas importantes atividades na Bioinformática são a genômica e a proteômica. A genômica se refere ao sequenciamento e análise de genomas, que abrange o conjunto completo de sequências de DNA (Ácido desoxirribonucleico) que codificam o material hereditário passado de geração para geração. Essas sequências de DNA incluem todos os genes e transcritos. Já a proteômica trata da análise do conjunto completo de proteínas, ou proteoma, que os genes expressam (FOX, 2005).

2.2 CONCEITOS BIOLÓGICOS

O material genético de todo ser vivo está codificado em longas cadeias de uma molécula chamada DNA. O DNA tem a forma de uma dupla hélice e é constituído por cadeias de ácidos nucleicos chamados nucleotídeos. No DNA são encontrados quatro tipos de nucleotídeos: Adenina (A), Citosina (C), Guanina (G) e Timina (T). Usualmente os nucleotídeos são chamados de bases, um segmento de uma molécula de DNA que contenha a informação requerida para a síntese de um produto biológico funcional, proteína ou RNA (Ácido Ribonucleico), é referido como um gene (NELSON; COX, 2006).

O RNA difere-se do DNA por uma discreta modificação no componente açúcar e pela presença da base Uracila (U) no lugar da Timina (T). Enquanto o principal papel do DNA é armazenar o código genético, o RNA apresenta diferentes tipos e funções. Os principais tipos de RNA são: o RNA mensageiro (mRNA), o RNA transportador (tRNA) e o RNA ribossômico (rRNA). O mRNA é transcrito a partir do DNA, e transporta a informação gênica até uma organela presente no citoplasma chamada ribossomo, responsável pela síntese de proteínas. O tRNA é responsável pelo transporte de moléculas de aminoácidos até o ribossomo. O rRNA é o principal componente dos ribossomos, responsável por catalisar algumas etapas no processo de síntese de proteínas (NELSON; COX, 2006).

O Dogma Central da Biologia Molecular é o conceito que descreve a maneira com que a informação gênica é transferida através de um sistema biológico. O conceito classifica em transferências gerais aquelas que representam o fluxo normal da informação biológica: o DNA pode ser copiado para DNA (Replicação), a informação do DNA pode ser copiada para um mRNA (Transcrição) e as proteínas podem ser sintetizadas a partir da informação no mRNA (Tradução).

Dada a existência de 20 aminoácidos e apenas quatro bases, o código genético é determinado por tríplexes de nucleotídeos (códon) que codificam um determinado aminoácido. Para cada códon presente no mRNA os tRNA associam um aminoácido gerando a sequência da proteína, sendo que a maioria dos aminoácidos é codificada por mais de um códon. O final da tradução é indicado por um códon de parada (*Stop codon*) que separa o mRNA do ribossomo liberando a proteína (WATSON et al., 2006). A TABELA 1 mostra os códon e os aminoácidos codificados.

TABELA 1 - CODIFICAÇÃO DE AMINOÁCIDOS

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

FONTE: (NASCIMENTO, 2005)

As regiões que codificam uma proteína de um mRNA são compostas por uma sucessão contínua de códon não sobrepostos, chamada fase aberta de leitura ou ORF (*Open Reading Frame*). Cada ORF define uma única proteína e tem início e fim localizados internamente ao mRNA (WATSON et al., 2006).

2.2.1 Anotação

A anotação de genomas consiste em obter informações estruturais e funcionais sobre uma ou várias sequencias relacionadas a um determinado genoma (STEIN, 2001). Para isso, é preciso integrar análises computacionais, dados biológicos auxiliares e perícia biológica para obter a maior quantidade possível de dados úteis (LEWIS; ASHBURNER; REESE, 2000).

Diversas ferramentas computacionais foram desenvolvidas com finalidades bem específicas como por exemplo identificação de genes, comparação entre sequencias e pesquisa em bases de dados. Essas três tarefas formam as bases para a anotação de genomas. As informações provenientes de anotações realizadas em milhares de organismos estão contidas em bancos de dados públicos como o GenBank, e são uma importante fonte de comparação entre sequencias genéticas.

2.3 ARMAZENAMENTO E OBTENÇÃO DE INFORMAÇÕES BIOLÓGICAS

2.3.1 FASTA

O formato de arquivos mais comum para representar sequencias biológicas, sejam elas de nucleotídeos ou resíduos de aminoácidos, é o formato FASTA. Tendo origem junto ao software homônimo (LIPMAN; PEARSON, 1985), tornou-se um padrão em bioinformática. No formato FASTA cada sequencia apresenta uma descrição indicada pelo símbolo "maior que" (>) ou por ponto e vírgula (;) no início da linha. A descrição não deve ultrapassar uma linha, caso ocorra, estas serão ignoradas pelo software servindo somente como comentário. Na linha seguinte à descrição tem-se a sequencia propriamente dita, sendo que pode haver quebra de linha. Um arquivo FASTA pode apresentar varias sequencias em um mesmo arquivo, sendo chamado então de Multi-FASTA onde cada sequencia apresenta a mesma formatação definida para sequencias únicas. A FIGURA 1 ilustra um exemplo de sequencia no formato FASTA.


```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLV
EWINGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILILLLLLLLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFPLIAGX
IENY
```

FIGURA 1 - FORMATO FASTA

FONTE: Wikipédia (2013)

2.3.2 GenBank

O GenBank é um banco de dados público criado e distribuído pelo *National Center for Biotechnology Information* (NCBI). O NR (*non-redundant*) é o maior banco não redundante de sequências de proteínas disponibilizado pelo NCBI, agregando dados de tradução de CDS (*Coding DNA Sequence*) do GenBank, dados dos bancos PDB (*Protein Data Bank*), SwissProt, PIR (*Protein Information Resource*) e PRF (*Protein Research Foundation*). Os dados do NCBI são disponibilizados gratuitamente através de arquivos de texto em formatos padronizados. Cada entrada possui uma descrição concisa com: nome científico, taxonomia, referências bibliográficas, lista de características, traduções em proteínas, unidades de transcrição, regiões repetitivas e sítios de mutação e modificação. Os arquivos seguem o padrão definido junto ao EMBL (*European Molecular Biology Laboratory*) que recebe o nome de GenBank (GBK), esse formato de arquivo descreve com detalhamento as características das sequências, um exemplo de registro pode ser visualizado na FIGURA 2.

```

LOCUS       LISOD                      756 bp    DNA        linear    BCT 30-JUN-1993
DEFINITION  Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION   X64011 S78972
VERSION     X64011.1  GI:44010
KEYWORDS    sod gene; superoxide dismutase.
SOURCE      Listeria ivanovii
  ORGANISM  Listeria ivanovii
            Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Haas,A. and Goebel,W.
  TITLE     Cloning of a superoxide dismutase gene from Listeria ivanovii by
            functional complementation in Escherichia coli and characterization
            of the gene product
  JOURNAL   Mol. Gen. Genet. 231 (2), 313-322 (1992)
  MEDLINE   92140371
REFERENCE   2  (bases 1 to 756)
  AUTHORS   Kreft,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1992) J. Kreft, Institut f. Mikrobiologie,
            Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG

FEATURES             Location/Qualifiers
     source           1..756
                     /organism="Listeria ivanovii"
                     /strain="ATCC 19119"
                     /db_xref="taxon:1638"
                     /mol_type="genomic DNA"
     RBS              95..100
                     /gene="sod"
     gene             95..746
                     /gene="sod"
     CDS              109..717
                     /gene="sod"
                     /EC_number="1.15.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="superoxide dismutase"
                     /db_xref="GI:44011"
                     /db_xref="GOA: P28763"
                     /db_xref="InterPro:IPR001189"
                     /db_xref="UniProtKB/Swiss-Prot:P28763"
                     /protein_id="CAA45406.1"
                     /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVS
                     GHAE LASKPGEELVANLDSVP E EIRGAVRNHGGGHANHTLFWSSLS PNGGGAPTGNLK
                     AAIESEFGTFDEFKEKFNAAAAARFGSGWAWLVVNNNGKLEIVSTANQDSPLSEGKT PV
                     LGLDVWEHAYYLKFKQNR RPEYIDTFWNVINWDERNKRFDAAK"
     terminator       723..746
                     /gene="sod"

ORIGIN
      1 cggtattttaa ggtgttacat agttctatgg aaatagggtc tatacctttc gccttacaat
     61 gtaatttctt .....
//

```

FIGURA 2 - EXEMPLO DE REGISTRO NO FORMATO GENBANK

FONTE: NCBI (2013)

2.3.3 COG

O banco de dados COG (*Clusters of Orthologous Groups*) (TATUSOV et al., 2003) classifica filogeneticamente os genes presentes em genomas bacterianos completos. Cada COG consiste em proteínas individuais ou grupos de parálogos com pelo menos três linhagens que representam um domínio conservado. A classificação se dá a níveis de categorias funcionais, como mostra a TABELA 2.

TABELA 2 - CATEGORIAS FUNCIONAIS DO COG

Código COG	Descrição
Processamento e armazenamento de informação	
J	Tradução, estrutura ribossomal e biogênese
A	Transformação e modificação de RNA
K	Transcrição
L	Replicação, recombinação e reparação
B	Estrutura e dinâmica da cromatina
Processos celulares e sinalização	
D	Controle do ciclo celular, divisão celular, e particionamento do cromossoma
Y	Estrutura nuclear
V	Mecanismos de defesa
T	Mecanismos de transdução de sinal
M	Biogênese da parede/membrana celular
N	Motilidade celular
Z	Citoesqueleto
W	Estruturas Extracelular
U	Tráfico intracelular, secreção e transporte vesicular
O	Modificação pós-traducional, renovação de proteínas, e chaperonas
Metabolismo	
C	Produção e conversão de energia
G	Transporte e metabolismo de carboidratos
E	Transporte e metabolismo de aminoácidos
F	Transporte e metabolismo de nucleotídeos
H	Transporte e metabolismo de Coenzima
I	Transporte e metabolismo de lipídios
P	Transporte e metabolismo de íons inorgânicos
Q	Biossíntese, transporte e catabolismo de metabólitos secundários
Não Caracterizadas	
R	Função geral predita
S	Função não conhecida

FONTE: (TIEPPO, 2011)

2.3.4 Pfam

O Pfam (FINN et al., 2010) é um banco de dados que contém uma vasta coleção de famílias de proteínas, cada uma representada por alinhamentos múltiplos e modelos escondidos de Markov (HMMs). Existem duas divisões nos bancos do Pfam: o Pfam-A,

que possui registros de alta qualidade curados manualmente e o Pfam-B, que possui dados gerados automaticamente com dados do banco ADDA (HEGER; HOLM, 2003).

O Pfam também fornece agrupamentos de alto nível, chamados clãs. Cada clã é uma coleção de registros do Pfam-A que possuem relação de similaridade, estrutura ou perfil de HMM.

2.3.5 ExPASy ENZYME

O servidor de dados proteômicos ExPASy (acrônimo para *Expert Protein Analysis System*) do *Swiss Institute of Bioinformatics* (SIB) é dedicado à análise de sequências e estruturas proteicas (GASTEIGER, 2003). O ENZYME é um dos bancos de dados agregados ao ExPASy, sendo um repositório de informações relacionadas à nomenclatura de enzimas.

Suas informações são baseadas nas recomendações do *Nomenclature Committee of the International Union of Biochemistry and Molecular Biology* (IUBMB) e contém para tipo de enzima:

- EC (*Enzyme Commission*) number;
- Nome recomendado;
- Nomes alternativos;
- Atividade catalítica;
- Cofatores;
- Referências para o SwissProt;
- Referências para o PROSITE (PROSITE, 2013);
- Referências para doenças humanas.

O banco de dados ENZYME pode auxiliar no desenvolvimento de aplicações que envolvem caminhos metabólicos, tendo em vista que tais bancos de dados descrevem coleções de enzimas, reações e caminhos bioquímicos que são utilizados por programas de computador para busca e visualização das informações metabólicas (BAIROCH, 2000).

2.4 SILA

A ferramenta de anotação automática SILA é fruto de um projeto de mestrado desenvolvido pelo aluno Ricardo Assunção Vialle do programa de pós-graduação em

bioinformática da UFPR. O sistema aqui proposto trabalha integrado ao SILA, fornecendo interface para a entrada de dados e trabalhando com os resultados fornecidos.

No SILA, as sequências são buscadas em três bancos de dados de proteínas utilizando a ferramenta RGS (*Rapid Gene Searcher*). As ORFs são comparadas com sequências de proteínas já anotadas presentes nos bancos de dados. Para que possa ser inferida a homologia a partir da similaridade.

As informações que podem ser agregadas à anotação são obtidas em três consultas independentes à três bancos de dados. A anotação final é apresentada no formato de arquivo GenBank (GBK) que possui campos específicos previstos para certas características. Na anotação, cada campo é preenchido caso obtenha um *hit* significativo no banco de dados. A FIGURA 3 exibe um exemplo de informações que podem ser associadas a um determinado gene.

- O campo Product é obtido pela descrição no banco NR do NCBI.
- O campo Function é obtido pelos grupos do COG.
- O campo EC Number é obtido pela descrição no banco Pfam.

```
/gene="0001.0008"
/function="[NOU]Cell motility|Posttranslational modification, protein turnover,
chaperones|Intracellular trafficking, secretion, and vesicular transport"
/product="Type 4 prepilin-like proteins leader peptide processing enzyme"
/EC_number="3.4.23.43"
/note="Organism:Methyloversatilis universalis FAM5 ScoreAlign:0.55301
ScoreRede:0.99837 ScoreCOG:0.52859 Query_coverage:91.32% Subject_coverage:92.28%"
/color=5
```

FIGURA 3 - EXEMPLO DE INFORMAÇÕES ASSOCIADAS NA ANOTAÇÃO

FONTE: Os autores (2013)

2.5 JSP

JavaServer Pages (JSP) é uma tecnologia baseada na linguagem de programação Java utilizada para o desenvolvimento de aplicações para Web que permite a execução em diversos sistemas operacionais. Com ela é possível que ao desenvolvedor criar aplicações que acessem o banco de dados, manipulem arquivos no formato texto, capturem informações a partir de formulários e captem informações sobre o visitante e sobre o servidor.

As aplicações JSP são carregadas e processadas dentro do servidor, permitindo que o código Java e certas ações pré-definidas sejam intercaladas com conteúdo de

marcação estático, resultando na página sendo compilada e executada no servidor para ser entregue num documento HTML (*HyperText Markup Language*) ou XML (*Extensible Markup Language*). As páginas compiladas e qualquer dependência de biblioteca Java devem ser executados com uma máquina virtual Java (JVM).

2.6 HTML5

HTML5 (*Hypertext Markup Language*, versão 5) é uma linguagem para estruturação e apresentação de conteúdo para a *World Wide Web* recomendada pela W3C (*World Wide Web Consortium*). Atualmente diversos navegadores já implementam algumas de suas funcionalidades. Ela traz consigo novas funcionalidades como semântica e acessibilidade. Sua essência tem sido melhorar a linguagem com o suporte para as mais recentes multimídias, enquanto a mantém facilmente legível.

O HTML5 adiciona novas funções sintáticas tais como o elemento <canvas>. Estas funções são projetadas para tornar mais fácil a inclusão e a manipulação de conteúdo gráfico e multimídia na web sem ter de recorrer a plugins proprietários e APIs.

Utilizamos neste trabalho duas bibliotecas específicas em HTML5, o Scribl (MILLER et al., 2013) para visualização de genomas e o RGraph (RGraph, 2013) para geração de gráficos.

2.6.1 Scribl

O Scribl é uma biblioteca para visualização genômica que utiliza as tecnologias do HTML5. Sendo voltada para a utilização com dados baseados em coordenadas, permite a visualização e a interação do usuário com elementos gráficos processados inteiramente no navegador. A visualização genômica apresenta opções de customização de objetos como por exemplo de modelos de genes, e da suporte para diversos formatos de dados como GenBank e BAM (MILLER et al., 2013).

2.6.2 RGraph

O RGraph, desenvolvido por Richard Heyes, é uma biblioteca para gráficos que utiliza JavaScript e HTML5. O RGraph apresenta diversos tipos de visualização de dados com mais de 20 gráficos disponíveis.

2.7 jQuery

jQuery é uma biblioteca de JavaScript desenvolvida com o objetivo de simplificar de tarefas executadas no lado do cliente em uma aplicação web. Sendo uma biblioteca livre de código aberto, diversos plugins com funcionalidades mais específicas foram desenvolvidas com o propósito de expandir as suas opções. Neste trabalho, utilizamos a biblioteca DataTables (DataTables, 2013) para fazer o tratamento das tabelas com paginação, ordenação e busca.

2.8 UML

A *Unified Modeling Language* (UML) é uma linguagem padronizada para a modelagem de sistemas orientados a objetos de modelagem não proprietária. A UML possui foco na especificação, documentação, estruturação e visualização lógica do desenvolvimento completo de um sistema.

A UML permite que desenvolvedores visualizem os produtos de seus trabalhos em diagramas padronizados. Junto com uma notação gráfica, a UML também especifica a semântica.

Utilizamos neste trabalho os diagramas de caso de uso, diagramas de classe, diagramas de atividades, diagramas de transição de estados, diagramas de componentes e diagramas de sequencia. O diagrama de caso de uso permite organizar os requisitos do sistema dando uma visão geral do sistema. O diagrama de classes representa a estruturação do sistema orientado a objetos. O diagrama de atividades representa os fluxos conduzidos por processamentos. O diagrama de transição de estados representa o estado ou situação em que os objetos podem se encontrar durante a execução do sistema. O diagrama de componentes ilustra os componentes envolvidos com as classes do sistema. O diagrama de sequencia serve uma visão sequencial das funcionalidades do sistema, baseado nas especificações de caso de uso.

2.9 MATLAB

MATLAB (*MATrix LABoratory*) é uma linguagem de programação e um ambiente de desenvolvimento desenvolvido pela MathWorks. O MATLAB permite a análise numérica,

manipulação de matrizes, processamento de sinais e construção de gráficos, além de fornecer interface para outras linguagens como C, C++, Java e Fortran.

A ferramenta SILA foi desenvolvida em MATLAB, dessa forma alguns elementos de integração, como o interpretador de arquivos GenBank, utilizados neste trabalho foram desenvolvidos em MATLAB.

3 METODOLOGIA

3.1 MODELO DE PROCESSO DE ENGENHARIA DE SOFTWARE

A metodologia utilizada para o desenvolvimento do projeto foi o modelo cascata. Essa metodologia utiliza uma abordagem sistemática, sequencial do desenvolvimento de software que se inicia no nível de sistema e avança ao longo da análise, projeto, codificação, teste e manutenção (PRESSMAN, 2005).

Este modelo de desenvolvimento foi escolhido para este trabalho devido as características bem definidas do sistema exigido. Já que na fase de análise de requisitos foram determinadas todas as funcionalidades necessárias e a maneira de realiza-las.

Segundo (PRESSMAN, 2005), o modelo cascata abrange as seguintes atividades:

- Análise e engenharia de sistemas: Estabelece os requisitos para todos os elementos do sistema e prossegue com a atribuição de partes dos requisitos ao sistema. Essa visão proporciona uma visão do sistema sobre a interface entre software, hardware, pessoas e bancos de dados.
- Análise de requisitos de software: Possui uma coleta de requisitos intensiva. Busca compreender o domínio da informação para o software, bem como a função, desempenho e interface exigidos. Esses requisitos são documentados e revistos com o cliente.
- Projeto: Envolve os passos de estrutura de dados, arquitetura de software, detalhes procedimentais e caracterização de interface. O projeto traduz as exigências numa representação que pode ser avaliada antes da codificação.
- Codificação: Traduz em forma legível para a máquina as instruções contidas no projeto.
- Testes: Os processos de teste concentram-se nos aspectos lógicos internos do software, garantindo que todas as instruções sejam testadas. Os teste também analisam os aspectos externos, garantindo que a entrada de dados produza resultados reais esperados.
- Manutenção: Reaplica cada uma das etapas precedentes do ciclo de vida ao programa existente.

Segue abaixo a divisão dos estágios no projeto do sistema. A divisão mais detalhada está disponível no apêndice (GRÁFICO DE GANTT).

- Estágio 1: Definição de escopo do projeto;
- Estágio 2: Levantamento de requisitos;
- Estágio 3: Definição das ferramentas utilizadas;
- Estágio 4: Desenvolvimento da integração com o software de anotação SILA;
- Estágio 5: Definição do layout das páginas;
- Estágio 6: Gerenciador de usuários;
- Estágio 7: Gerenciador de tarefas;
- Estágio 8: Exibição da anotação com o visualizador Scribl;
- Estágio 9: Realizar formas de agrupamento dos dados;
- Estágio 10: Exibição dos agrupamentos com RGraph;
- Estágio 11: Exibição das informações específicas em tabelas;
- Estágio 12: Revisão e finalização do software.

A definição do escopo e o levantamento de requisitos se basearam em sistemas semelhantes já disponíveis. Definimos quais informações podiam ser extraídas das anotações geradas pelo SILA e como mostra-las de maneira interativa proporcionando fácil acesso.

O desenvolvimento da integração dos software SILA com o sistema online através de banco de dados no estágio 4 exigiu esforço da equipe devido a necessidade de transpor as informações antes contidas em arquivos texto para um modelo relacional de banco de dados, foi utilizado o SGBD MySQL. O desenvolvimento de gráficos de visualização também exigiram esforços devido a adaptação da equipe com as funcionalidades do HTML5. A biblioteca Scribl foi utilizada para exibir o mapa gráfico do genoma e a biblioteca RGraph foi utilizada para exibir os gráficos de agrupamento.

Os demais estágios foram menos complexos. A equipe conseguiu realizar essas tarefas em pouco tempo, devido a familiaridade com os processos necessários envolvidos.

3.2 PLANO DE ATIVIDADES

As atividades foram divididas de acordo com a demanda a cada estágio do projeto. Procurou-se separar as tarefas específicas para cada membro, de forma que cada um pudesse executá-las de forma independente. Assim, efetuou-se uma média de tempo de dedicação semanal por membro da equipe. As etapas e tarefa desenvolvidas no projeto estão especificadas nos apêndices PGP – PLANO DE GERENCIAMENTO DE PROJETO - GRÁFICO DE GANNT, DIVISÃO DE TRABALHO NO PROJETO.

3.3 PLANO DE RISCOS

O plano de riscos está descrito no apêndice no apêndice PGP – PLANO DE GERENCIAMENTO DE PROJETO - RISCOS DO PROJETO.

3.4 RESPONSABILIDADES

As responsabilidades do projeto estão especificadas no apêndice PGP – PLANO DE GERENCIAMENTO DE PROJETO - ORGANIZAÇÃO DO PESSOAL.

3.5 MATERIAIS

As características de hardware e software utilizados para o desenvolvimento do projeto estão especificadas no apêndice PGP – PLANO DE GERENCIAMENTO DE PROJETO - RECURSOS DO PROJETO.

3.6 DESENVOLVIMENTO DO PROJETO

A definição do escopo e das funcionalidades do sistema foram rápidas. Devido ao tema de bioinformática atendido pelo sistema, necessitou-se dos membros da equipe o entendimento dos conceitos biológicos envolvidos de forma a entender quais as possibilidades poderiam exploradas.

A escolha da linguagem JSP decorreu devido a robustez e das ferramentas disponíveis para a integração com o software SILA, além da experiência de todos os membros com a linguagem. A escolha do HTML5 ocorreu devido a nova tendência de

padrão web que vem sendo seguida no mercado, atendendo as necessidades de visualização de gráficos interativos antes atendidos por tecnologias proprietários como o FLEX.

A divisão das tarefas foram feitas com escopos bem definidos para membro, de maneira que o andamento das partes do projeto atendidas por cada membro não fosse afetado por outro. Utilizamos o repositório virtual e controlador de versão Assembla para integrar o desenvolvimento entre a equipe. A equipe utilizou o ambiente de desenvolvimento Eclipse que possibilita integração com o repositório direto pela interface.

4 APRESENTAÇÃO DO SOFTWARE

4.1 ACESSO AO SISTEMA

A tela inicial (FIGURA 4FIGURA 4) do sistema apresenta a opção para usuários já cadastrados acessarem as funcionalidades através do login. É oferecido também um link para realizar novo cadastro (FIGURA 5FIGURA 5).

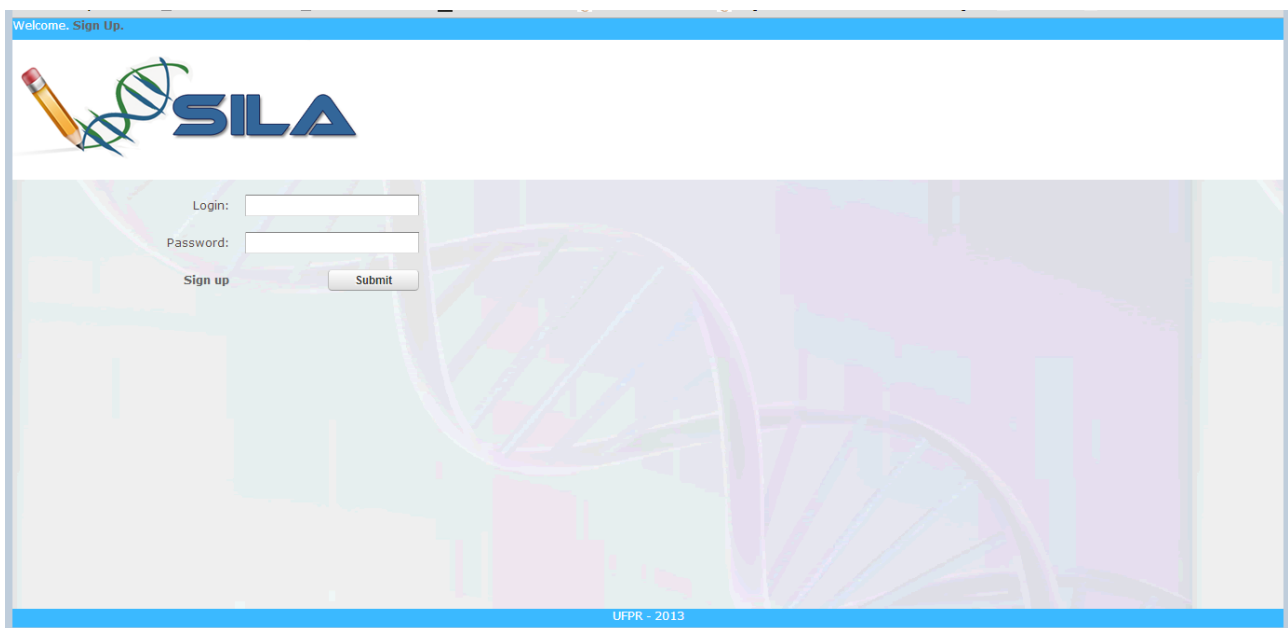
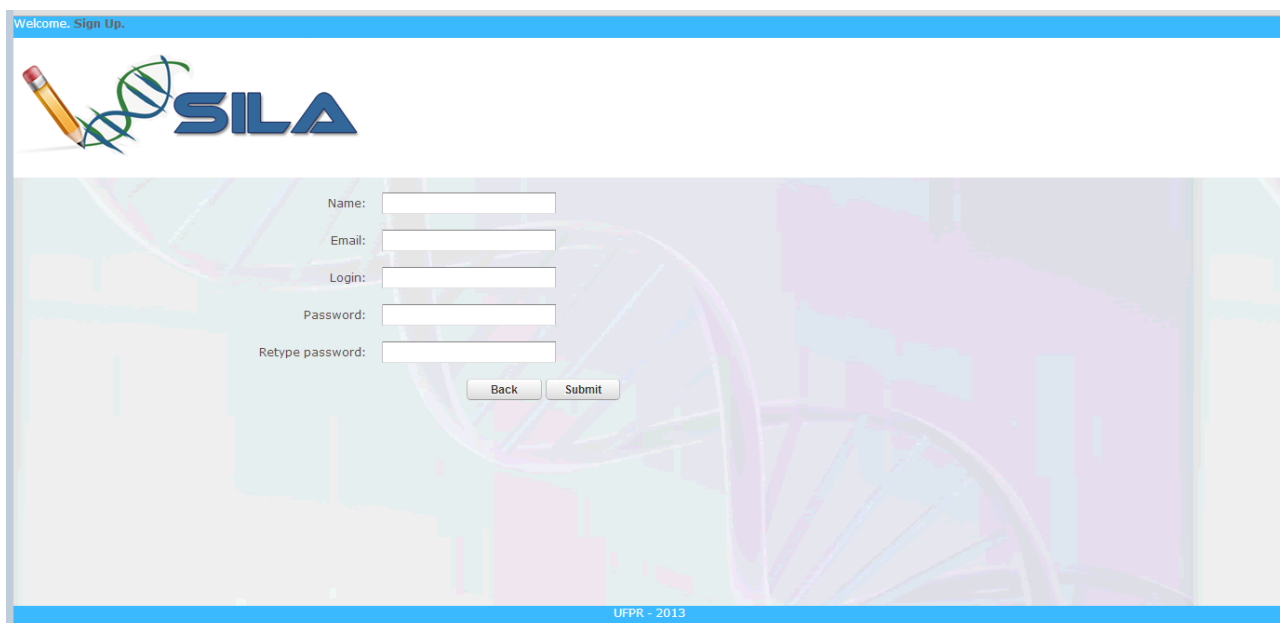



FIGURA 4 - TELA DE LOGIN

FONTE: Os autores (2013)



Welcome. Sign Up.



Name:

Email:

Login:

Password:

Retype password:

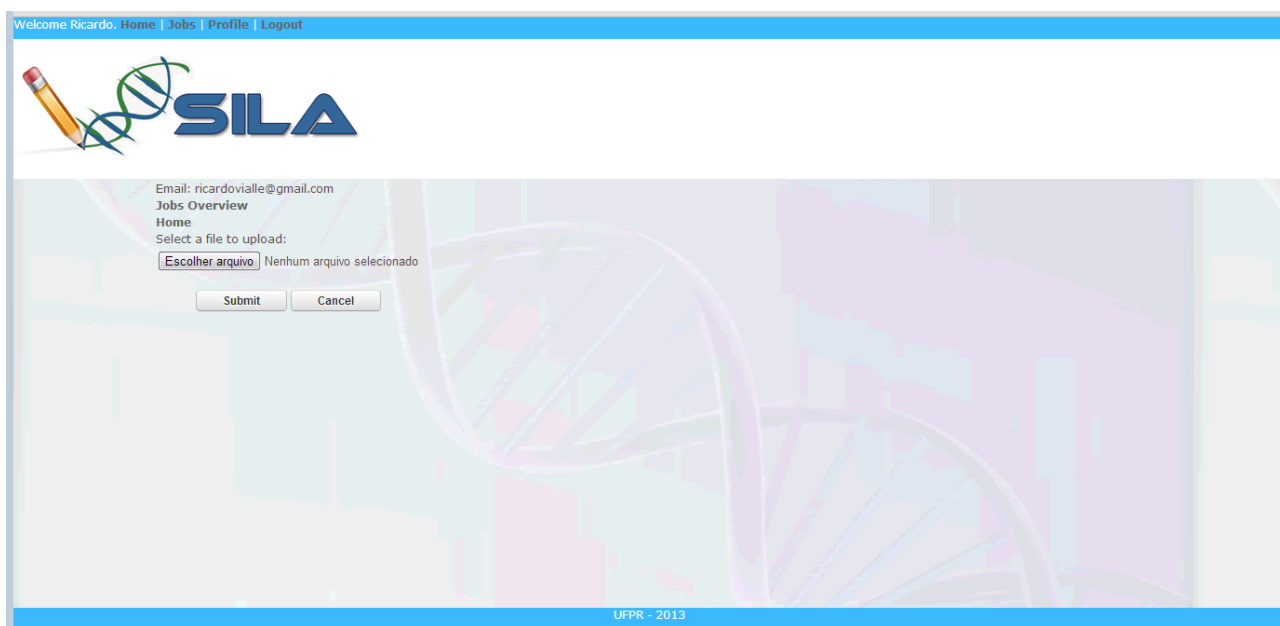
UFPR - 2013

FIGURA 5 - TELA DE CADASTRO


FONTE: Os autores (2013)

4.2 ENVIAR ARQUIVO PARA ANOTAÇÃO

Após efetuado o login, o usuário terá acesso a tela principal onde é disponibilizada a opção para submissão de um novo arquivo no formato FASTA (FIGURA 6). Além disso o usuário pode ter acesso as informações de tarefas submetidas e alteração nas informações pessoais.



Welcome Ricardo. Home | Jobs | Profile | Logout



Email: ricardovialle@gmail.com

Jobs Overview

Home

Select a file to upload:

Nenhum arquivo selecionado

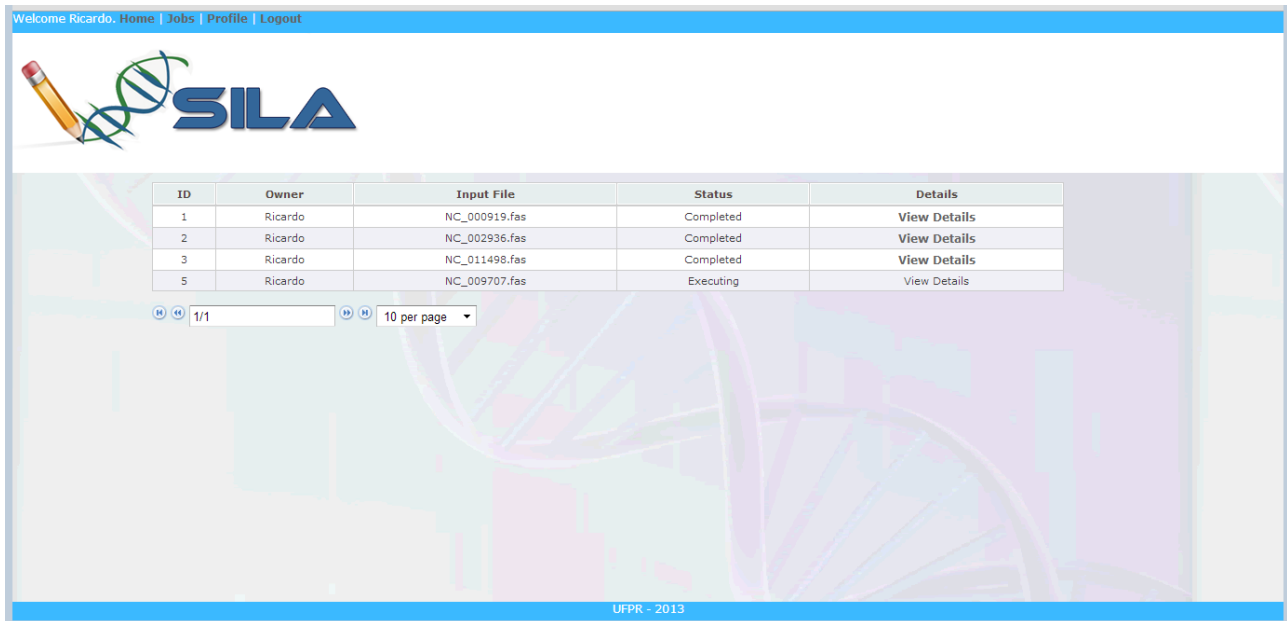
UFPR - 2013

FIGURA 6 - TELA PRINCIPAL

FONTE: Os autores (2013)

4.3 TAREFAS SUBMETIDAS

Ao submeter uma tarefa o usuário pode visualiza-las de forma tabelada no qual são fornecidas informações gerais da tarefa (FIGURA 7). Quando a tarefa apresenta status igual a "Completed" o usuário tem acesso a informações detalhadas da anotação.



ID	Owner	Input File	Status	Details
1	Ricardo	NC_000919.fas	Completed	View Details
2	Ricardo	NC_002936.fas	Completed	View Details
3	Ricardo	NC_011498.fas	Completed	View Details
5	Ricardo	NC_009707.fas	Executing	View Details

1/1 10 per page

UFPR - 2013

FIGURA 7 - TELA DE TAREFAS

FONTE: Os autores (2013)

4.4 DETALHES DA ANOTAÇÃO

Ao selecionar para visualizar os detalhes de uma anotação. A tela de detalhes (FIGURA 8) é exibida. Nessa tela são fornecidas:

- Opção para download do arquivo de anotação no formato GenBank;
- Visualização dos genes posicionalmente no genoma;
- Tabela com informações gerais de todos os genes;
- Links para visualização por agrupamento.

Informações detalhadas sobre cada gene podem ser acessadas tanto através do gráfico do genoma como pela tabela.

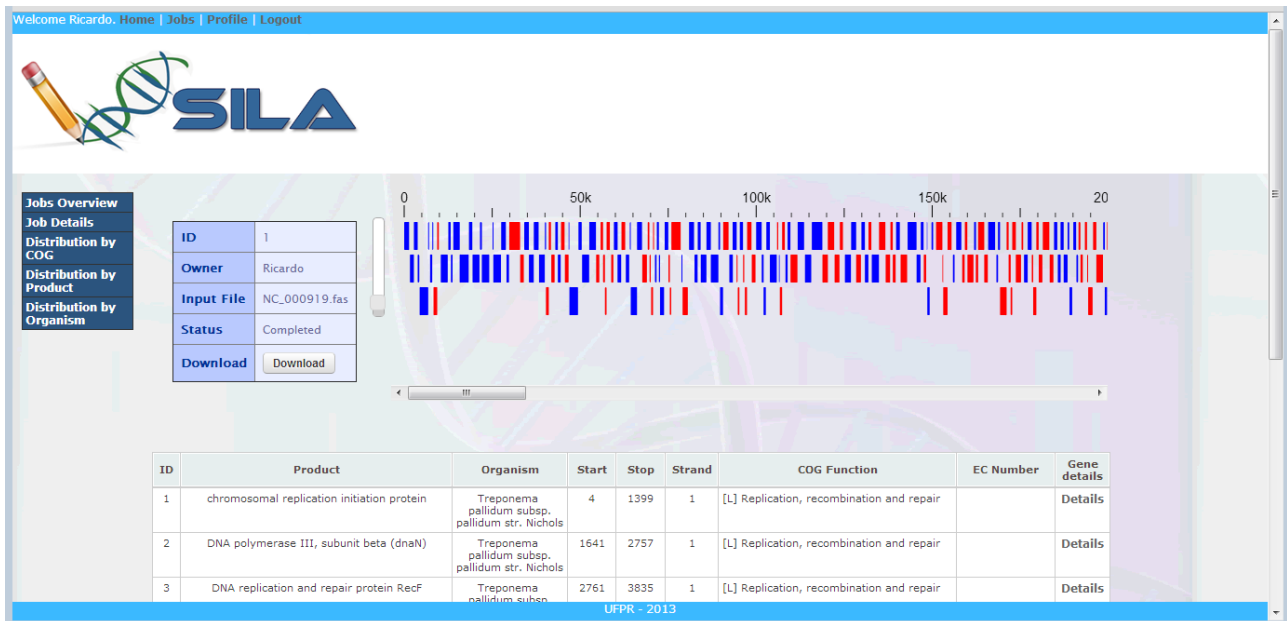


FIGURA 8 - TELA DE DETALHES DA ANOTAÇÃO

FONTE: Os autores (2013)

4.5 AGRUPAMENTOS

O sistema apresenta três formas de agrupamento dos genes:

- Por produto (FIGURA 10);
- Por organismo (FIGURA 11);
- Por função (COG) (FIGURA 9).

As informações são exibidas em gráficos e tabelas. Tanto nos agrupamentos por produto como por organismo, os gráficos exibem um conjunto limitado de categorias, ranqueadas pela quantidade de genes associados. No caso do COG que apresenta uma quantidade de classificações fixas, todas são exibidas nos gráficos. Nas tabelas não há esta distinção, todas as informações são disponibilizadas.

Os agrupamentos do COG e organismo apresentam dois níveis de visualização. No COG, primeiramente são exibidas as distribuições por grupos gerais que ao serem acessados apresentam as classes relacionadas. Na distribuição por organismo, primeiramente são exibidos os agrupamentos por gênero, e no segundo nível as espécies do gênero selecionado. As tabelas seguem a mesma lógica de níveis.

Após o ultimo nível de visualização, ao selecionar um agrupamento é exibida uma tabela com os genes referentes ao filtro realizado (FIGURA 12).

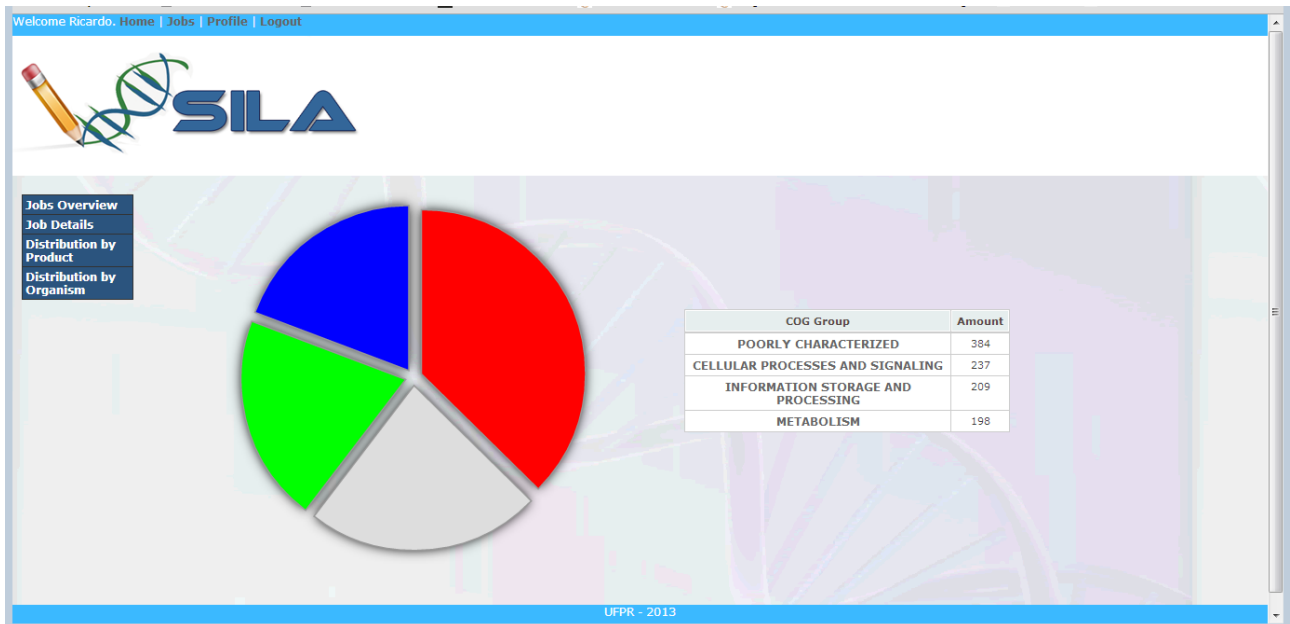


FIGURA 9 - TELA DISTRIBUIÇÃO POR COG

FONTE: Os autores (2013)

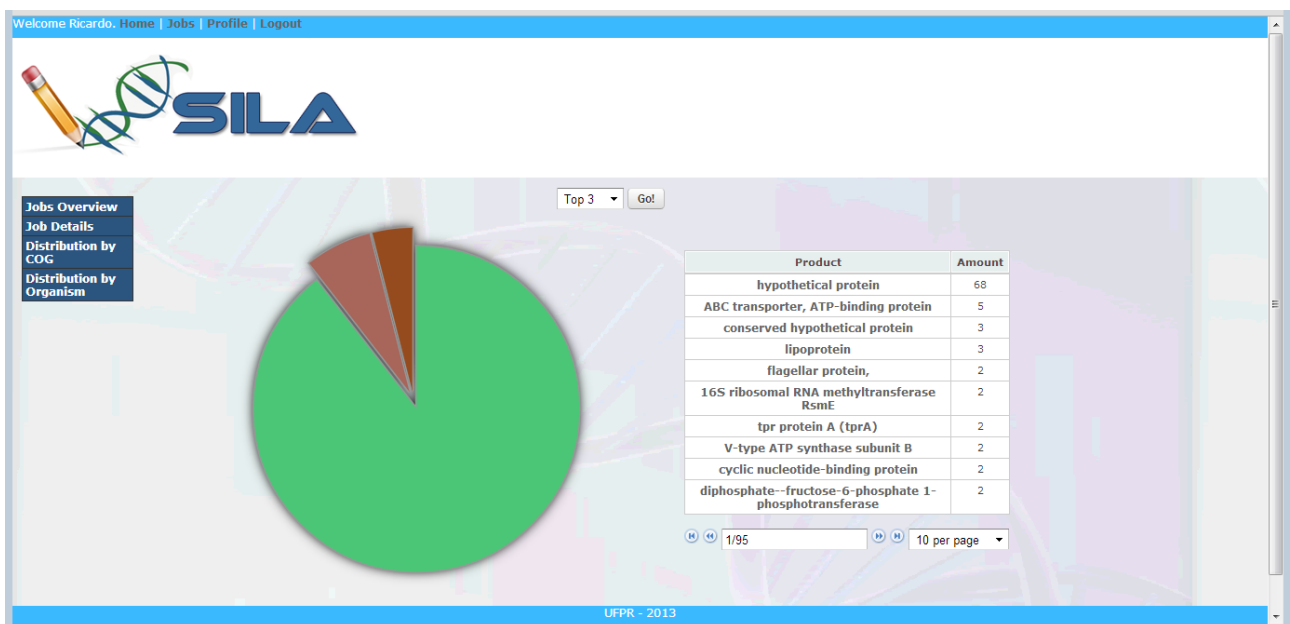


FIGURA 10 - TELA DISTRIBUIÇÃO POR PRODUTO

FONTE: Os autores (2013)

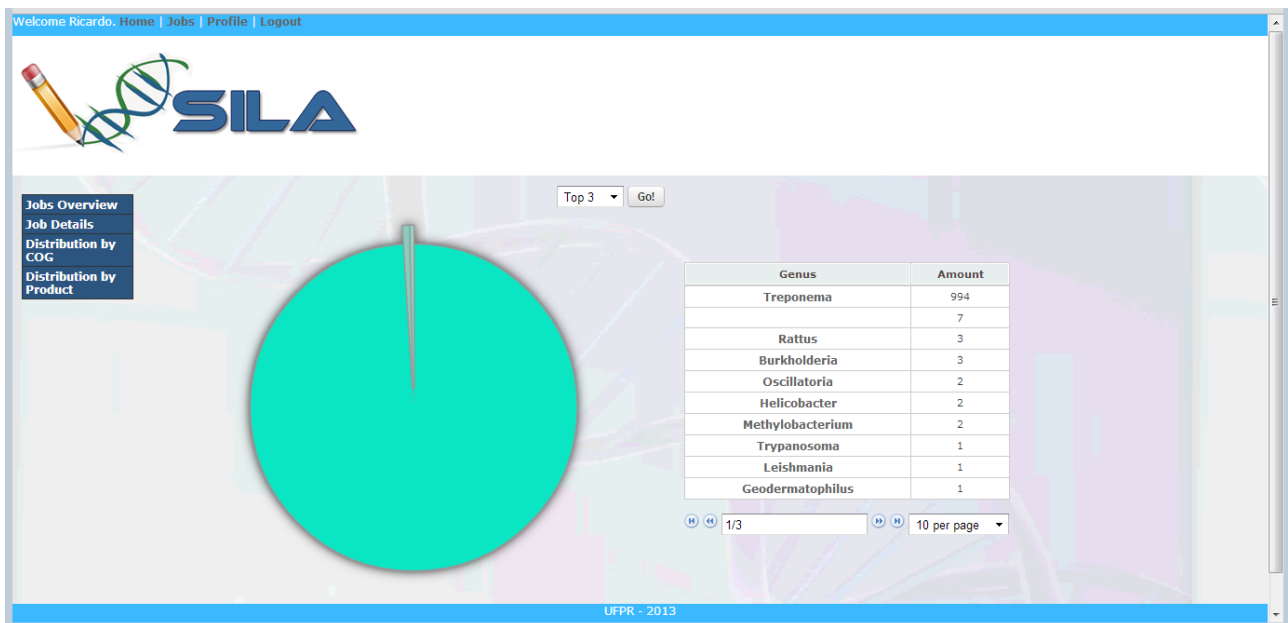


FIGURA 11 - TELA DISTRIBUIÇÃO POR ORGANISMO

FONTE: Os autores (2013)

Welcome Ricardo, Home | Jobs | Profile | Logout

SILA

Jobs Overview
Job Details
Distribution by Product
Distribution by Organism
Back

ID	Product	Organism	Start	Stop	Strand	COG Function	EC Number	Gene details
61	hypothetical protein	Treponema pallidum subsp. pallidum str. Chicago	71268	71703	1	[S] Function unknown		Details
66	hypothetical protein	Treponema pallidum subsp. pallidum DAL-1	75472	75094	-1	[S] Function unknown		Details
124	hypothetical protein	Treponema pallidum subsp. pertenue str. Gauthier	148283	147617	-1	[S] Function unknown		Details
125	hypothetical protein	Treponema pallidum subsp. pertenue str. SamoaD	148476	148602	1			Details
132	hypothetical protein	Treponema pallidum subsp. pallidum DAL-1	155885	156380	1	[S] Function unknown		Details
134	hypothetical protein	Treponema pallidum subsp. pallidum str. Chicago	159213	158538	-1	[S] Function unknown		Details
239	hypothetical protein	Treponema pallidum subsp. pallidum str. Chicago	254509	257833	1	[S] Function unknown		Details
276	hypothetical protein	Treponema pallidum subsp. pallidum str. Chicago	298495	298636	1			Details
289	hypothetical protein	Treponema pallidum subsp. pallidum str. Chicago	310225	310999	1	[S] Function unknown		Details
296	hypothetical protein	Treponema pallidum subsp. pallidum str. Chicago	320847	317931	-1	[S] Function unknown		Details

1/7 10 per page

UFPR - 2013

FIGURA 12 - TABELA DE GENES AGRUPADOS

FONTE: Os autores (2013)

4.6 DETALHES DOS GENES

As informações sobre cada gene são exibidas em tela separada (FIGURA 13). Nela são apresentadas todas as informações relacionadas pelo processo de anotação, exibindo inclusive a sequência de DNA referente ao gene.

Informações referente ao número EC podem ser acessadas também em tela separada (FIGURA 14), que fornece as informações contidas no banco de dados ExPASy ENZYME.

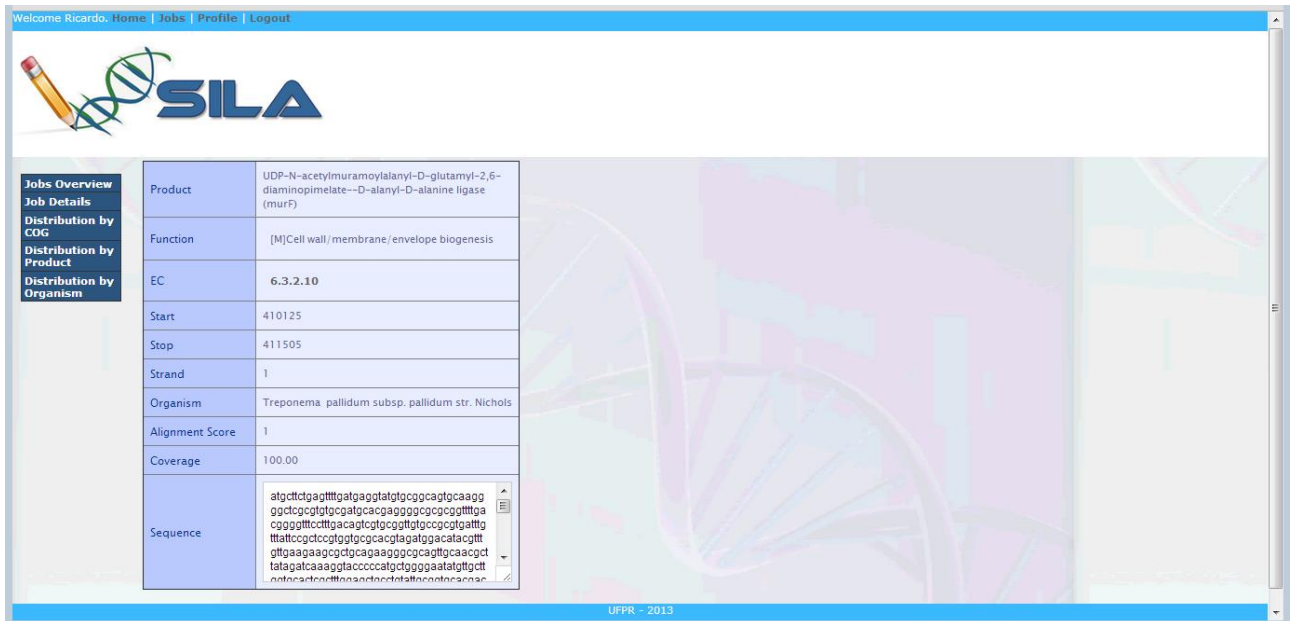


FIGURA 13 - TELA DETALHES DO GENE

FONTE: Os autores (2013)

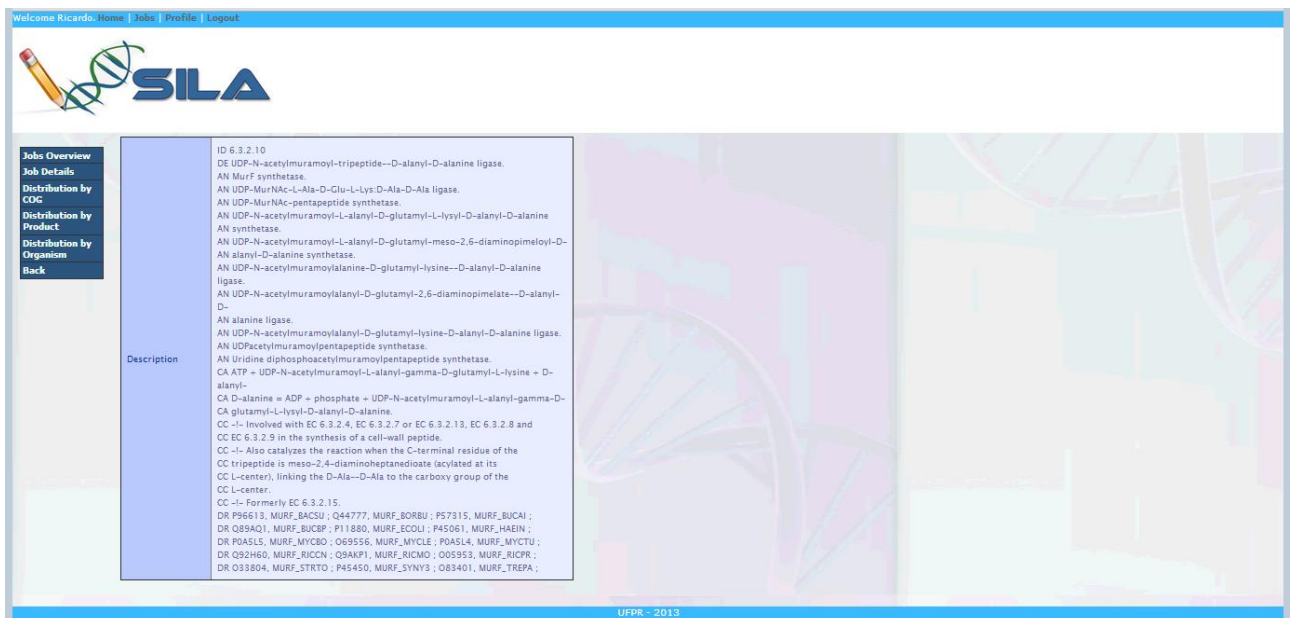


FIGURA 14 - TELA INFORMAÇÕES DA ENZIMA

FONTE: Os autores (2013)

5 CONSIDERAÇÕES FINAIS

O planejamento e desenvolvimento do sistema nos proporcionou um contato inicial e prático com a bioinformática. Ao realizar o projeto e trabalhar com o conceito de anotação de genomas e com os formatos de arquivos específicos utilizados foi possível entender uma parte dos problemas desta área.

Conseguimos criar uma ferramenta que integra as funcionalidades de uma ferramenta de anotação automática de genomas para o ambiente web. Criamos uma nova interface visual, baseada nas necessidades de visualização dos dados. O sistema teve foco na usabilidade e otimização para visualização dos dados com objetos gráficos.

As funcionalidade de gráficos necessitam de suporte à HTML5 por parte do navegador do usuário. Os navegadores mais usados em suas versões mais atuais já apresentam suporte as essas funcionalidades. Caso o navegador não possua suporte a HTML5, as informações ainda podem ser acessadas através das tabelas, não afetando a usabilidade do sistema.

Algumas situações realizam acesso direto ao arquivo de anotação GenBank. Para exibir a visualização gráfica do genoma e para exibir as sequencias dos genes, as informações do arquivo são acessadas via JavaScript. Dessa forma evitamos a inclusão de algumas informações no banco de dados que poderiam afetar a performance do sistema.

Para criação das tabelas de genes, criamos algumas buscas especificas para lidar com o grande volume de dados. A etapa de montagem de tabela é a mais custosa do sistema, principalmente quando visualizamos todos os genes sem filtro algum. Utilizamos métodos de paginação e ordenação em JavaScript que necessitam que toda a tabela seja carregada previamente, dessa forma há certa lentidão para carregar algumas páginas.

Durante o processo, várias ideias novas foram criadas, mas nem todas implementadas. Isso pelo fato do tempo de projeto, escopo e complexidade. Porém todas as ideias sugeridas foram analisadas. Com isto é possível uma ampliação do sistema, tornando mais abrangente e com mais funções agregadas. Algumas funcionalidades e melhorias que podem ser implementadas são listadas abaixo:

- Suporte à anotação manual: A funcionalidade que permite a alteração manual das anotações é algo comum nos sistemas de anotação, porém sua implementação foge do escopo do projeto.

- Integração com ferramentas de comparação de sequências: Confirmar as informações das sequências através de buscas remotas ao sistema do BLAST.
- Fornecer o serviço de anotação via *Web Service*: Muitas ferramentas de bioinformática trabalham integradas a serviços remotos, a implementação de um *Web Service* permite que a ferramenta possa ser agregada a outras ferramentas de forma remota.
- Fornecer outros formatos de anotação: Utilizamos um formato de anotação bastante completo, porém alguns outros formatos são também bastante utilizados, como por exemplo o GFF. A disponibilização de outros formatos para download é uma opção bem vinda.

REFERÊNCIAS

BAIROCH, A. The ENZYME database in 2000. **Nucleic acids research**, v. 28, n. 1, p. 304-5, 1 jan. 2000.

BAXEVANIS, A.; OUELLETTE, B. **Bioinformatics**. New York, USA: John Wiley & Sons, Inc., 2001. v. 43

DataTables. Disponível em: <<http://www.datatables.net/>>. Acesso em: 19 jan. 2013.

FINN, R. D. et al. The Pfam protein families database. **Nucleic acids research**, v. 38, n. Database issue, p. D211-22, jan. 2010.

FOX, J. **WHAT IS BIOINFORMATICS?** Disponível em: <<http://www.scq.ubc.ca/what-is-bioinformatics/>>. Acesso em: 19 jan. 2013.

GASTEIGER, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, v. 31, n. 13, p. 3784-3788, 1 jul. 2003.

HEGER, A.; HOLM, L. Exhaustive enumeration of protein domain families. **Journal of molecular biology**, v. 328, n. 3, p. 749-67, 2 maio. 2003.

LESK, A. M. **Introduction to bioinformatics**. New York, USA: Oxford University Press, 2002.

LEWIS, S.; ASHBURNER, M.; REESE, M. G. Annotating eukaryote genomes. **Current opinion in structural biology**, v. 10, n. 3, p. 349-54, jun. 2000.

LIPMAN, D.; PEARSON, W. Rapid and sensitive protein similarity searches. **Science**, v. 227, n. 4693, p. 1435-1441, 22 mar. 1985.

MILLER, C. A et al. Scribl: an HTML5 Canvas-based graphics library for visualizing genomic data over the web. **Bioinformatics (Oxford, England)**, v. 29, n. 3, p. 381-3, 1 fev. 2013.

NASCIMENTO, L. V. DO. **Um Sistema Baseado em Agentes para Re-anotação de Genomas**. [S.l.] Universidade Federal do Rio Grande do Sul, 2005.

NCBI. Disponível em: <<http://www.ncbi.nlm.nih.gov/>>. Acesso em: 18 jan. 2013.

NELSON, D. L.; COX, M. M. **Lehninger princípios de bioquímica**. 4. ed. São Paulo: Sarvier, 2006.

PRESSMAN, R. S. **Engenharia de software**. Terceira ed. São Paulo: Pearson Education, 2005.

PROSITE. Disponível em: <<http://prosite.expasy.org/>>. Acesso em: 1 fev. 2013.

RGraph. Disponível em: <<http://www.rgraph.net/>>. Acesso em: 5 jan. 2013.

ROUZÉ, P.; PAVY, N.; ROMBAUTS, S. Genome annotation: which tools do we have for it? **Current opinion in plant biology**, v. 2, n. 2, p. 90-5, abr. 1999.

STEIN, L. Genome annotation: from sequence to biology. **Nature reviews genetics**, v. 2, n. July, 2001.

TATUSOV, R. L. et al. The COG database: an updated version includes eukaryotes. **BMC bioinformatics**, v. 4, p. 41, 11 set. 2003.

TIEPPO, E. **MONTAGEM E ANÁLISE PRELIMINAR DO GENOMA DE *Bradyrhizobium elkanii* 587 UTILIZANDO LEITURAS DE SEQUÊNCIAS DE DNA CURTAS**. [S.l.] Universidade Federal do Paraná, 2011.

WATSON, J. D. et al. **Biologia molecular do gene**. 5. ed. Porto Alegre: Artmed, 2006. p. 760

APÊNDICES

1 DIAGRAMA DE CASO DE USO

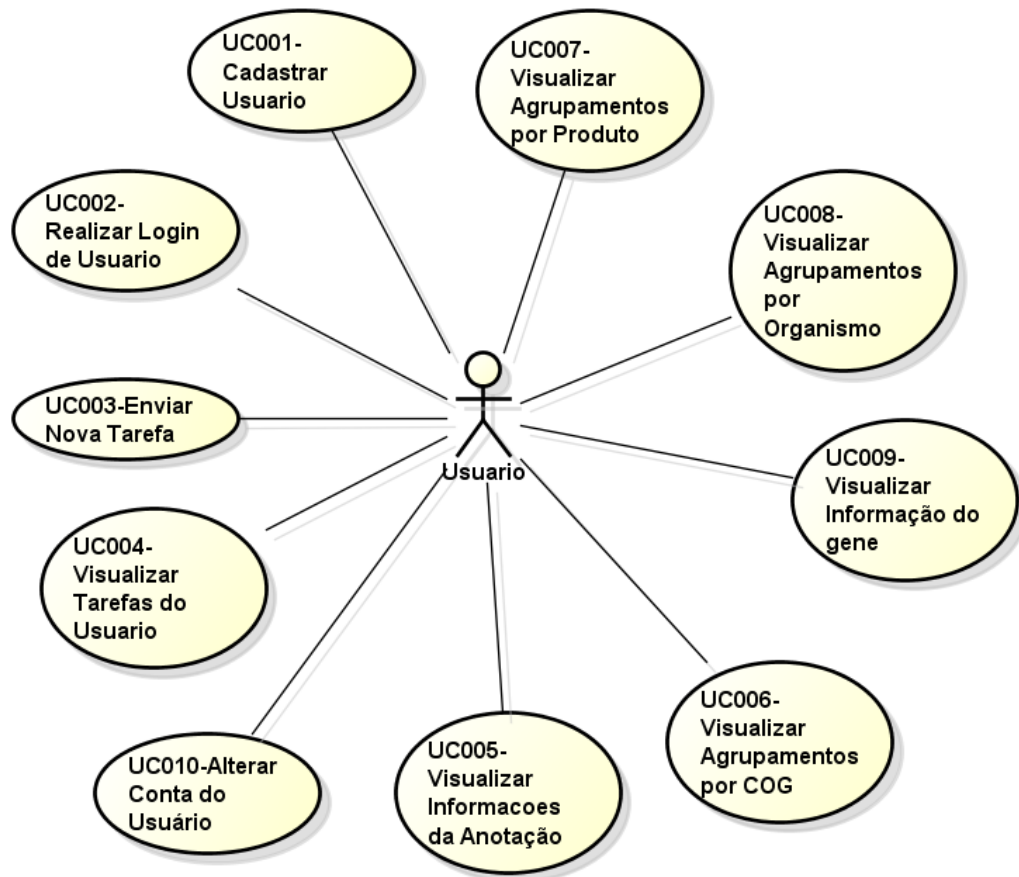


FIGURA 15 - DIAGRAMA DE CASOS DE USO

FONTE: Os autores (2013)

2 DIAGRAMA DE TRANSIÇÃO DE ESTADOS

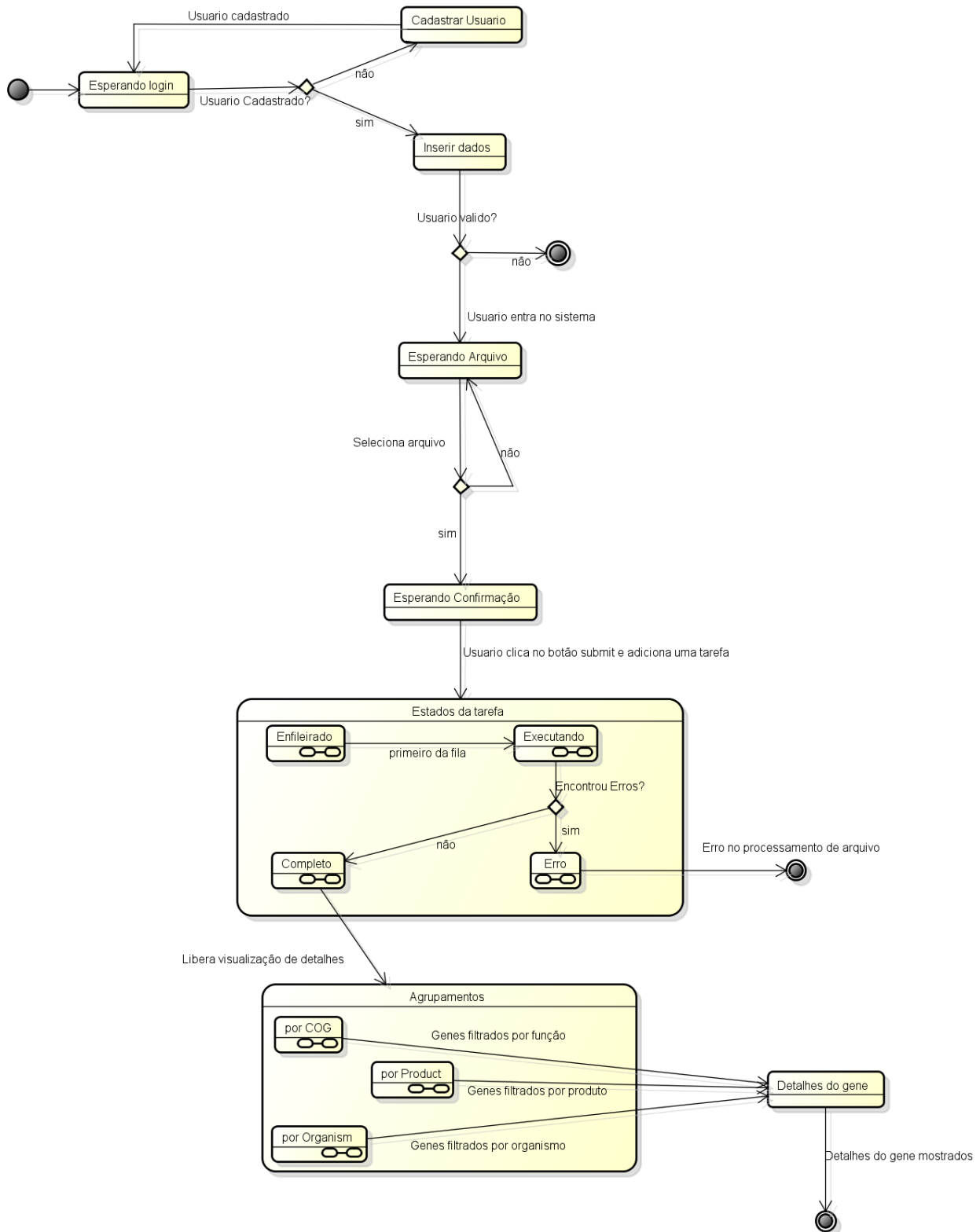


FIGURA 16 - DIAGRAMA DE TRANSIÇÃO DE ESTADOS

FONTE: Os autores (2013)

3 DIAGRAMA DE COMPONENTES

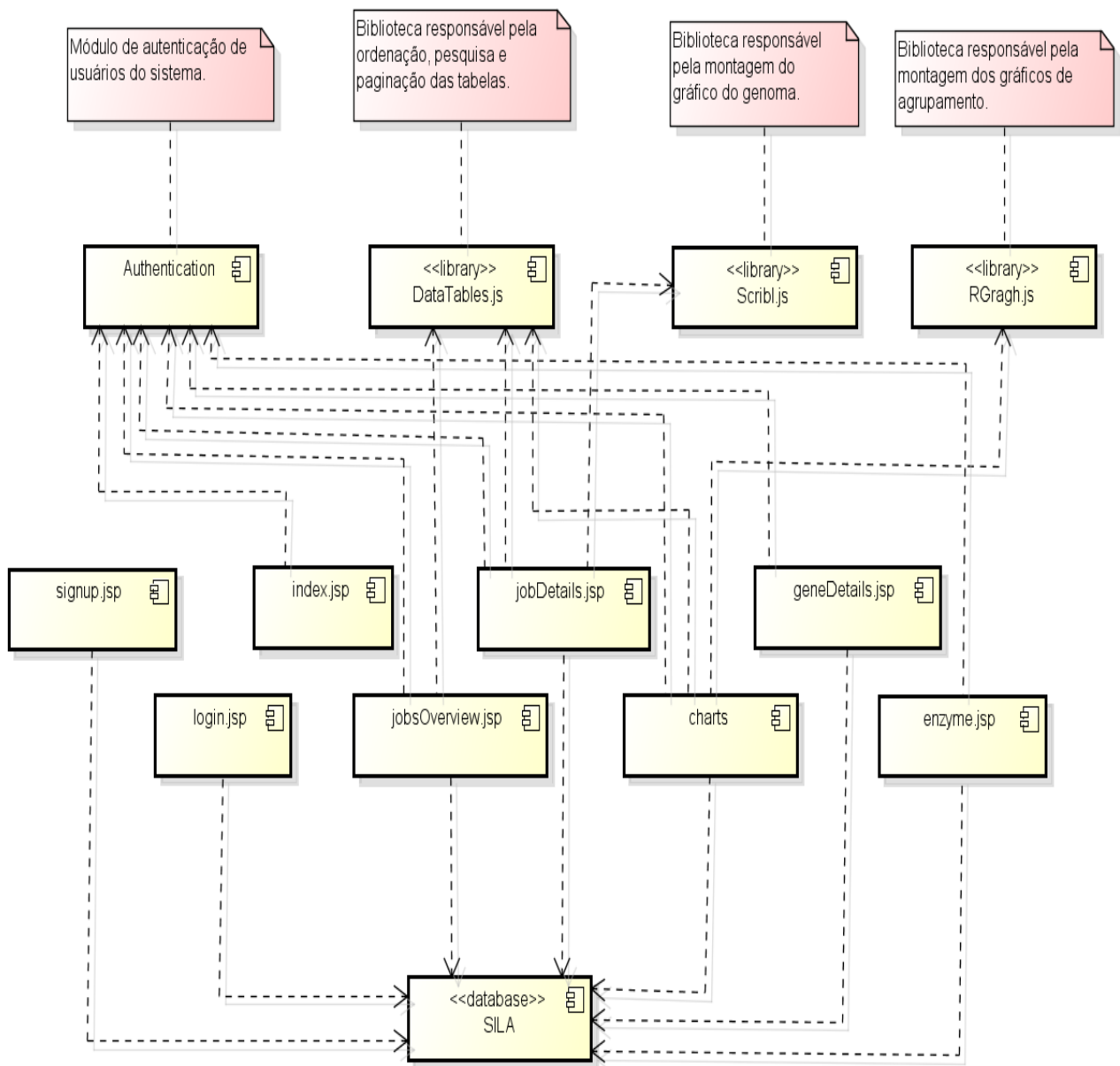


FIGURA 17 - DIAGRAMA DE COMPONENTES

FONTE: Os autores (2013)

4 DIAGRAMA DE ATIVIDADES

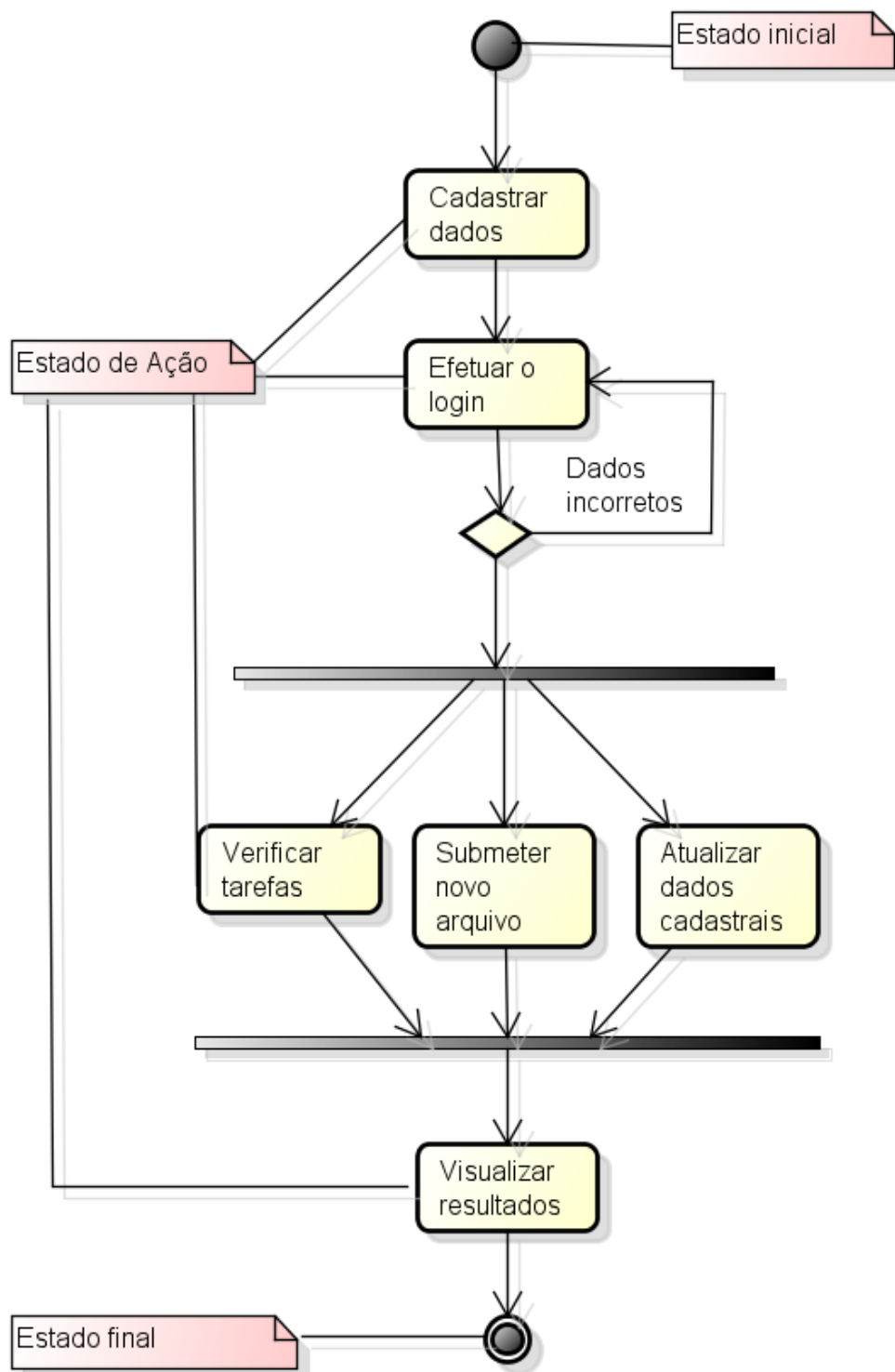


FIGURA 18 - DIAGRAMA DE ATIVIDADES

FONTE: Os autores (2013)

5 DIAGRAMA DE CLASSES

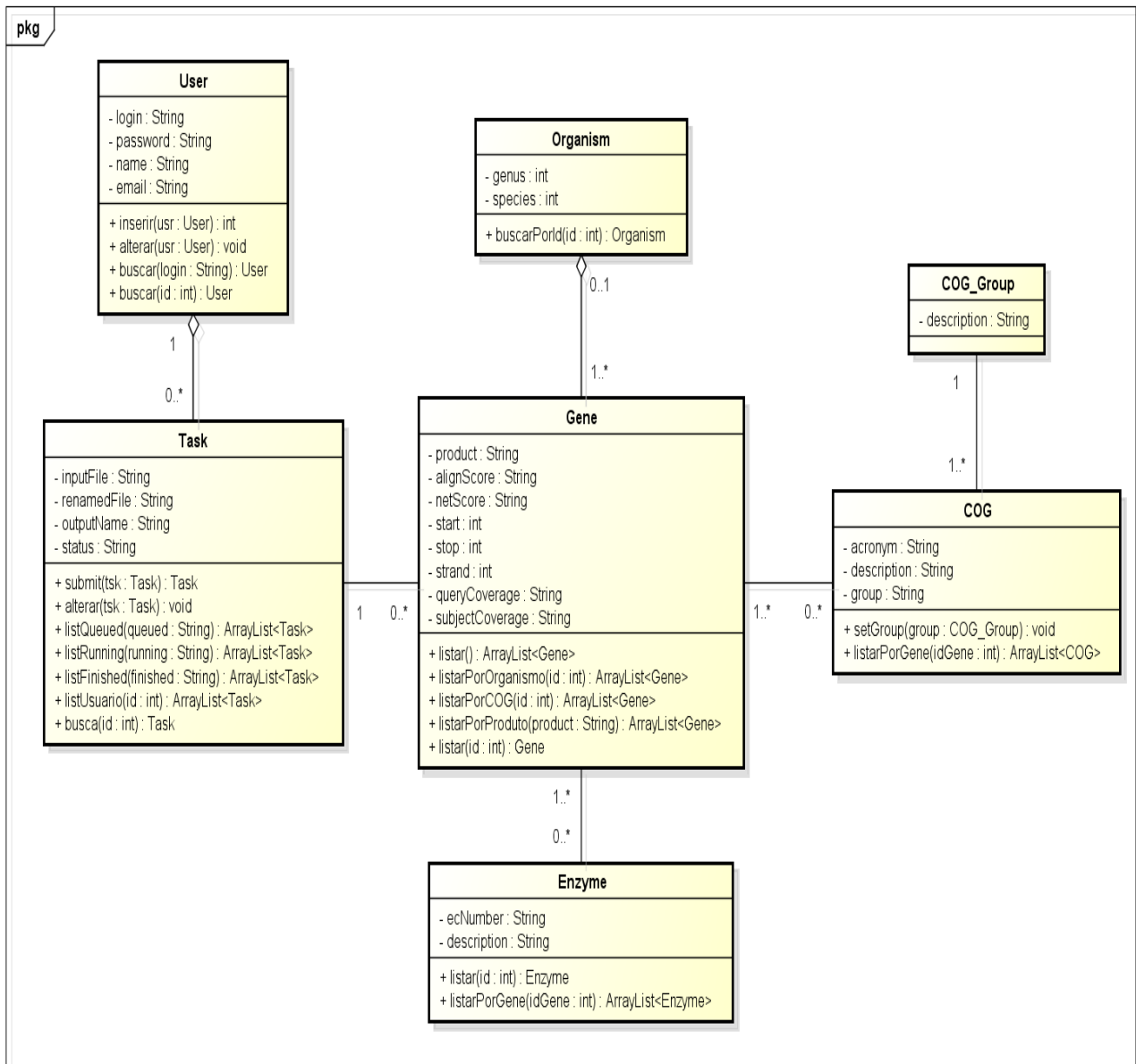


FIGURA 19 - DIAGRAMA DE CLASSES

FONTE: Os autores (2013)

6 DIAGRAMA ENTIDADE RELACIONAMENTO

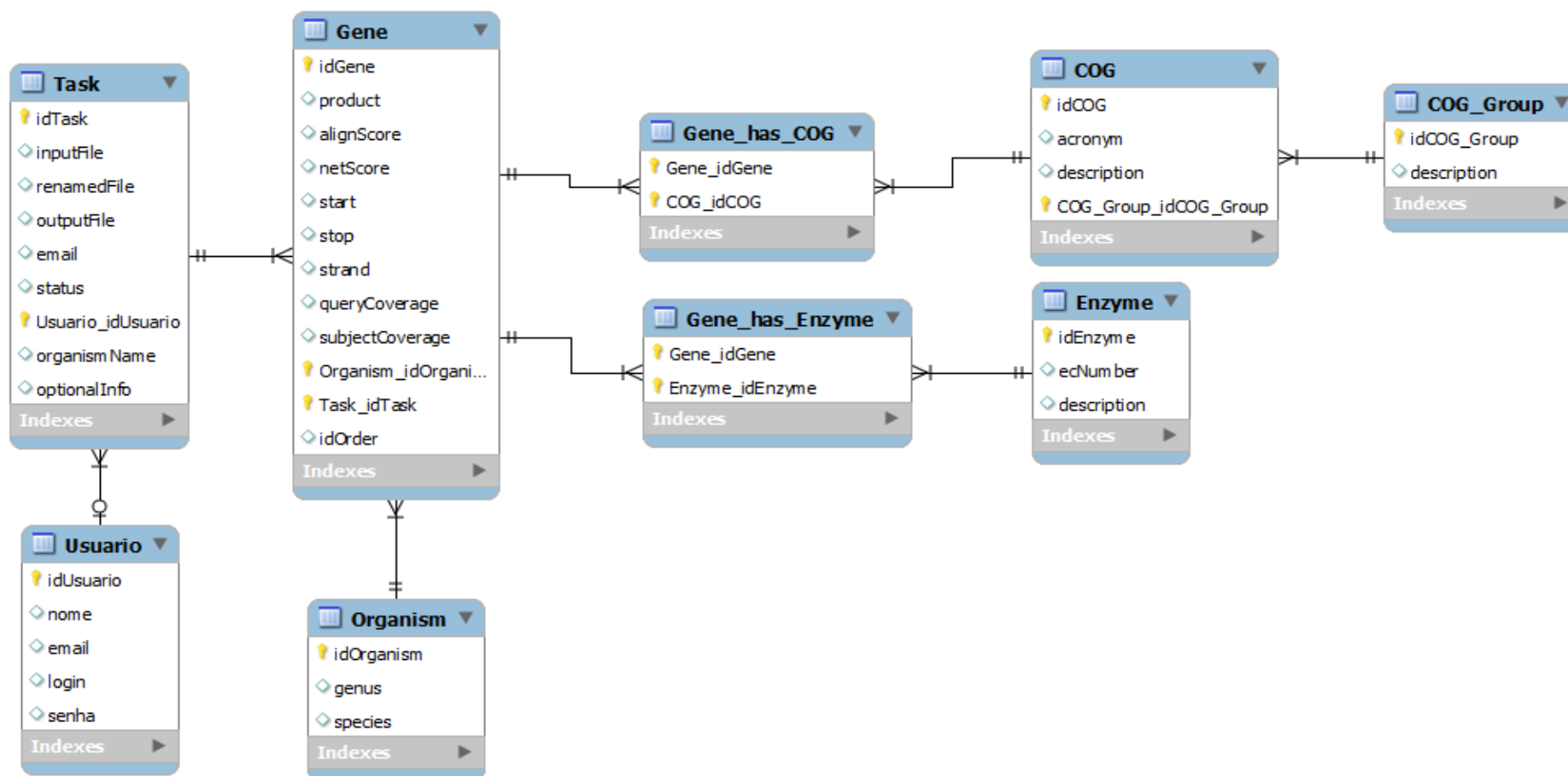


FIGURA 20 - DIAGRAMA ENTIDADE RELACIONAMENTO

FONTE: Os autores (2013)

7 ESPECIFICAÇÃO DE CASOS DE USO

7.1 UC001 – Cadastrar Usuário

Descrição

Este caso de uso serve para cadastrar dados de um usuário.

Data View

DV1 – Tela para cadastro de usuário

Welcome. [Sign Up.](#)

SILA

Name:

Email:

Login:

Password:

Retype password:

UFPR - 2013

Pré-condições

Não há pré-condições para este caso de uso.

Pós-condições

Após o fim normal deste caso de uso o sistema deve:

1. O usuário deverá conseguir executar o cadastro no sistema.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema apresenta a tela **(DV1)**
2. O usuário preenche o campo Name **(E1)(A1)**
3. O usuário preenche o campo Email **(R1)(E2)(E3)**
4. O usuário preenche o campo Login **(E4)**
5. O usuário preenche o campo Password **(E5)**
6. O usuário preenche o campo Retype password **(E6)(E7)**
7. O usuário pressiona o botão Submit **(A2)**
8. Sistema insere no banco os dados do novo usuário.
9. O sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
10. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Sign up clicado.

1. Sistema recarrega a página apresentada na tela **DV1**.

A2: Botão Back clicado.

1. Sistema chama o Use Case **UC002 – Realizar Login de Usuário**

Fluxos de Exceção

E1: Campo Name não preenchido.

1. Sistema apresenta a mensagem “Please provide a name” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E2: Campo Email não preenchido.

1. Sistema apresenta a mensagem “Please provide a valid email address” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E3: Campo Email preenchido incorretamente.

1. Sistema consiste os dados inseridos no campo Email
2. Sistema apresenta a mensagem “Please provide a valid email address” ao usuário.
3. O usuário pressiona o botão OK da mensagem.
4. Retorna ao fluxo principal.

E4: Campo Login não preenchido.

1. Sistema apresenta a mensagem “Please provide a login” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E5: Campo Password não preenchido.

1. Sistema apresenta a mensagem “Please provide a password” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E6: Campo Retype password não preenchido.

1. Sistema apresenta a mensagem “Please retype your password” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E7: Campo Password e Retype password preenchidos com valores diferentes.

1. Sistema apresenta a mensagem “The provided passwords are different.” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

Regras de Negócio

- R1.** O campo Email deve ser preenchido com um endereço de e-mail válido.

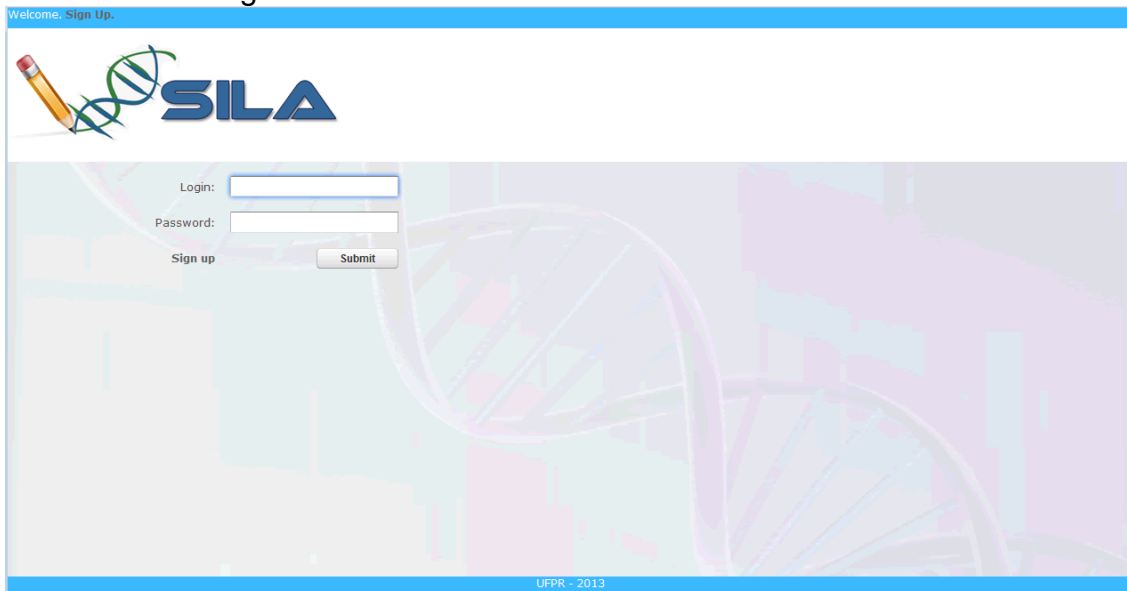
7.2 UC002 – Realizar Login de Usuário

Descrição

Este caso de uso serve para realizar o login do usuário.

Data View

DV1 – Tela de login de usuário.



Pré-condições

O usuário deverá estar cadastrado no sistema.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir executar o login no sistema.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema apresenta a tela **(DV1)**
2. O usuário preenche o campo Login **(A1)**
3. O usuário preenche o campo Password
4. O usuário pressiona o botão Submit **(E1)(E2)**
5. Sistema consiste os dados preenchidos pelo usuário.
6. Sistema chama o Use Case **UC001 – Cadastrar de Usuário.**
7. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Sign up clicado.

1. Sistema chama o Use Case **UC001 – Cadastrar de Usuário.**

Fluxos de Exceção

E1: Campo login não preenchido.

1. Sistema apresenta a mensagem "Please provide your login" ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E2: Campo password não preenchido.

1. Sistema apresenta a mensagem "Please provide your password" ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

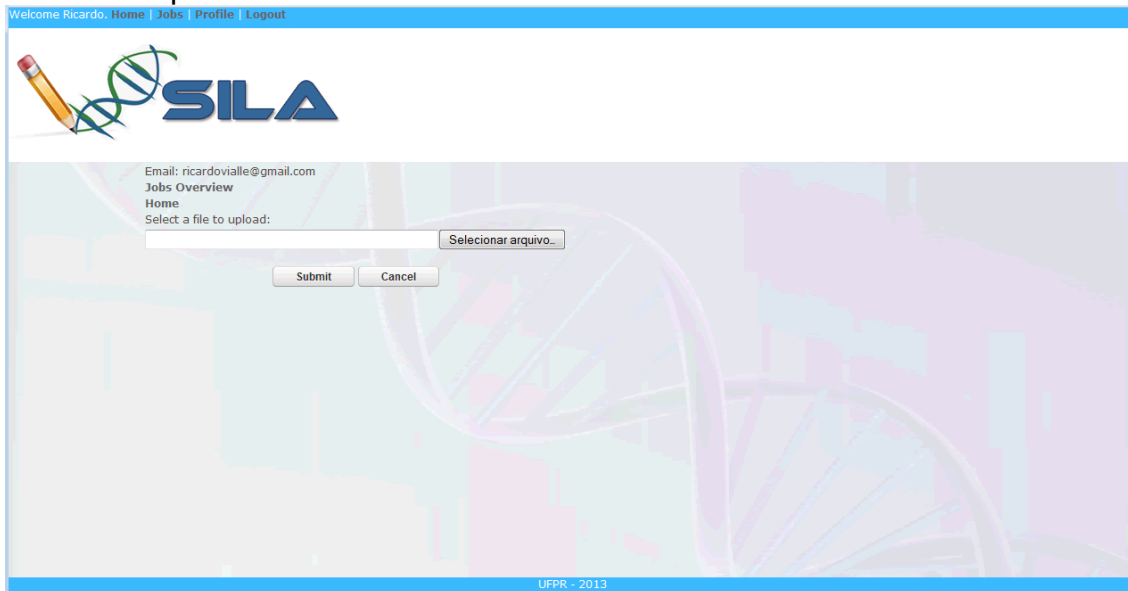
7.3 UC003 – Enviar Nova Tarefa

Descrição

Este caso de uso serve para o usuário submeter uma nova tarefa.

Data View

DV1 – Tela para enviar uma nova tarefa.



Pré-condições

O usuário deverá estar logado no sistema.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir submeter uma nova tarefa.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema busca o email do usuário que está logado.
2. O sistema apresenta a tela **(DV1)**
3. O usuário clica no botão para selecionar o arquivo **(E1)(A1)(A2)(A3)(A4)(A5)**
4. Usuário seleciona o arquivo correspondente a tarefa.
5. Sistema preenche campo texto com o caminho do arquivo selecionado.
6. O usuário pressiona o botão **Submit (A6)**
7. Sistema atualiza banco de dados e insere o novo arquivo na pasta uploads.
8. O sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
9. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. Caso de uso é finalizado.

A2: Link Jobs Overview clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**

2. Caso de uso é finalizado.

A3: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário**.
2. Caso de uso é finalizado.

A4: Link Profile clicado.

1. Caso de uso é reiniciado.

A5: Link Logout clicado.

1. Sistema efetua o logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário**.
3. Caso de uso é finalizado.

A6: Botão Cancel clicado.

1. Sistema limpa o campo texto.
2. Retorna ao passo 3 do fluxo principal.

Fluxos de Exceção

E1: Botão Submit clicado.

1. Sistema apresenta a mensagem “Please select a file” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

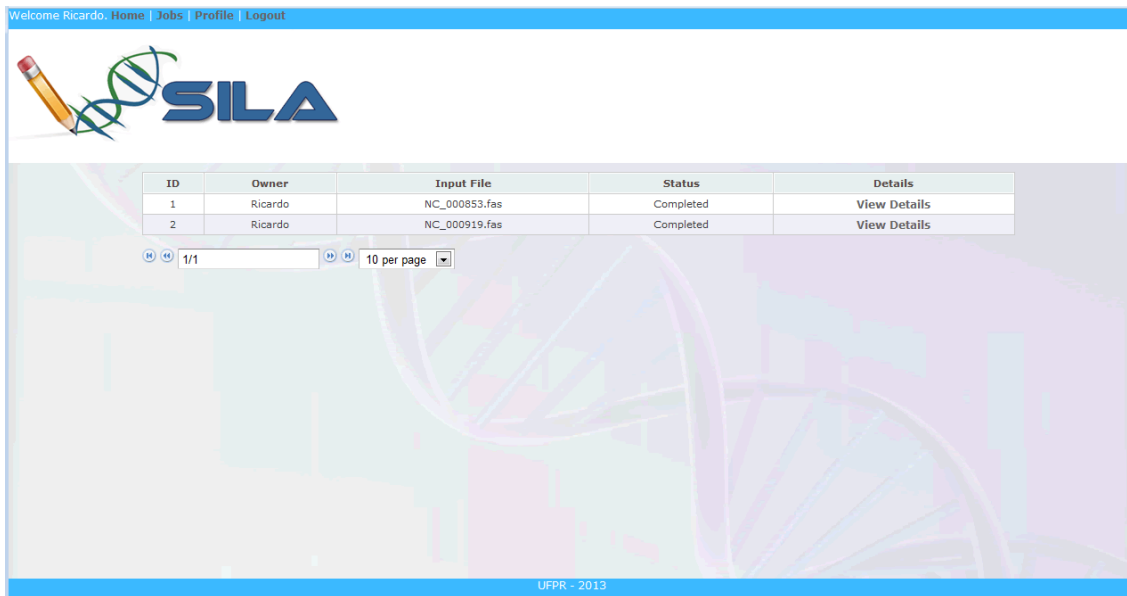
7.4 UC004 – Visualizar Tarefas do Usuário

Descrição

Este caso de uso serve para o usuário visualizar as suas tarefas enviadas.

Data View

DV1 – Tela de visualização das tarefas do usuário.



Pré-condições

O usuário deverá estar logado no sistema e ter clicado no link “Jobs Overview”.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir visualizar suas tarefas.

Ator primário

Sistema

Fluxo de Eventos Principal

1. O sistema faz a busca das tarefas. **(R1)**
2. O sistema preenche a tabela de tarefas com os dados retornados. **(R2)**
3. O sistema apresenta a tela **(DV1)**
4. O usuário visualiza suas tarefas **(A1) (A2) (A3) (A4) (A5)**
5. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa**.
2. O caso de uso é finalizado.

A2: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário**.
2. O caso de uso é finalizado.

A3: Link Profile clicado.

1. Sistema chama o Use Case **UC010 – Alterar Conta do Usuário**.
2. O caso de uso é finalizado.

A4: Link Logout clicado.

1. Sistema realiza logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário**.
3. O caso de uso é finalizado.

A5: Link View Details clicado.

1. Sistema chama o Use Case **UC005 – Visualizar Informações da Anotação**.

Regras de Negócio

R1: O sistema realiza uma busca das tarefas associadas ao usuário.

R2: Caso o sistema não encontre tarefas relacionadas ao usuário o sistema exibira a mensagem “No Jobs Found”.

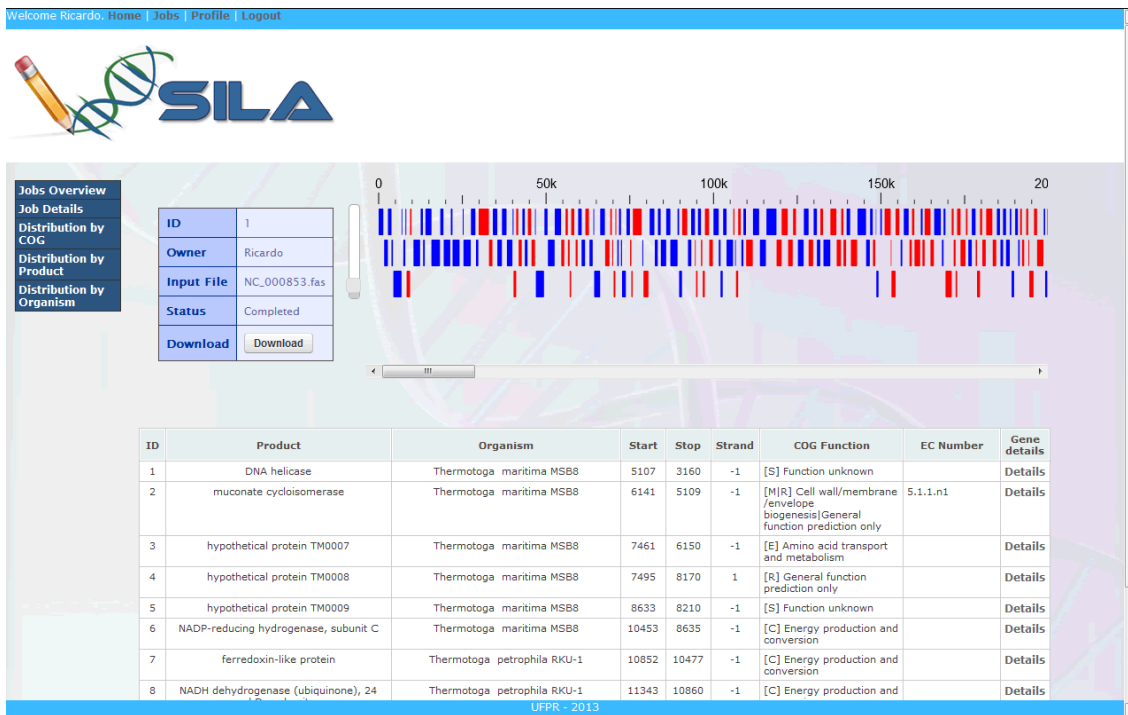
7.5 UC005 – Visualizar Informação da Anotação

Descrição

Este caso de uso serve para o usuário visualizar informações das suas tarefas enviadas.

Data View

DV1 – Tela de visualização das informações da tarefa do usuário.



Pré-condições

O usuário deverá estar logado no sistema e ter clicado no link “View Details”.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir visualizar informações gerais da anotação.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema faz a busca informações sobre o arquivo correspondente a sequencia genética. **(R1)**
2. O sistema monta o gráfico baseado nas informações buscadas.
3. O sistema preenche as tabelas com os dados da busca. **(R2)**
4. O sistema apresenta a tela **(DV1)**
5. O usuário visualiza as informações da anotação **(A1) (A2) (A3) (A4) (A5) (A6) (A7) (A8) (A9) (A10) (A11) (A12)**
6. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**

2. O caso de uso é finalizado.

A2: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário**.
2. O caso de uso é finalizado.

A3: Link Profile clicado.

1. Sistema chama o Use Case **UC010 – Alterar Conta do Usuário**.
2. O caso de uso é finalizado.

A4: Link Logout clicado.

1. Sistema realiza logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário**.
3. O caso de uso é finalizado.

A5: Botão Download pressionado.

1. Sistema fornece a função para download do arquivo GBK.
2. O caso de uso é finalizado.

A6: Link Jobs Overview clicado

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário**.
2. O caso de uso é finalizado.

A7: Link Jobs Details clicado

1. O caso de uso é reiniciado.

A8: Link Distribution by COG clicado.

1. Sistema chama o Use Case **UC006 – Visualizar Agrupamentos por COG**.
2. O caso de uso é finalizado.

A9: Link Distribution by Product clicado.

1. Sistema chama o Use Case **UC007 – Visualizar Agrupamentos por Product**.
2. O caso de uso é finalizado.

A10: Link Distribution by Organism clicado.

1. Sistema chama o Use Case **UC008 – Visualizar Agrupamentos por Organism**.
2. O caso de uso é finalizado.

A11: Link Gene no gráfico clicado.

1. Sistema chama o Use Case **UC009 – Visualizar Informações do Gene**.
2. O caso de uso é finalizado.

A12: Link Details clicado.

1. Sistema chama o Use Case **UC009 – Visualizar Informações do Gene**.
2. O caso de uso é finalizado.

Regras de Negócio

R1: O sistema realiza uma busca dos resultados da tarefa selecionada pelo usuário preenchendo e montando o gráfico com essas informações.

R2: O sistema realiza uma busca dos dados relacionados a tarefa e seus resultados, preenchendo a tabela lateral ao gráfico com os dados da tarefa e a tabela na parte inferior da pagina com os dados dos genes relacionados a tarefa.

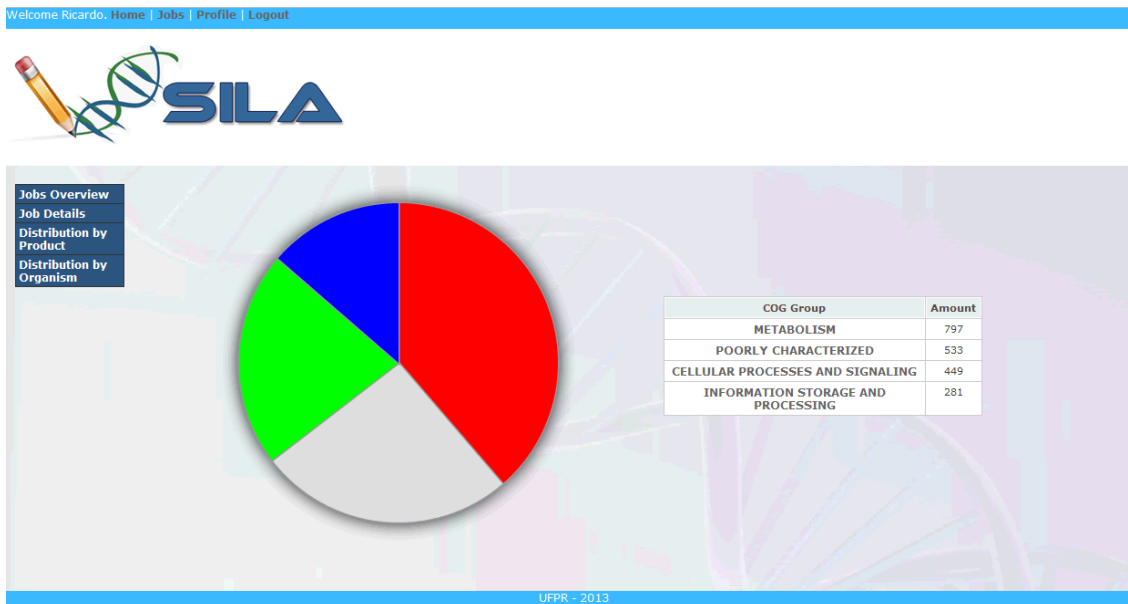
7.6 UC006 – Visualizar Agrupamentos por COG

Descrição

Este caso de uso serve para o usuário visualizar os agrupamentos por COG.

Data View

DV1 – Tela de visualização do gráfico de agrupamento por COG.



Pré-condições

O usuário deverá estar logado no sistema e ter clicado no link “Distribution by COG”.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir visualizar os agrupamentos por COG.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema faz a busca dos dados de agrupamentos por COG, baseado nos parâmetros passados via url. **(R1)**
2. O sistema monta o gráfico com base nos dados retornados.
3. O sistema monta as tabelas com base nos dados retornados.
4. O sistema apresenta a tela **(DV1)**
5. O usuário visualiza as informações desejadas **(A1) (A2) (A3) (A4) (A5) (A6) (A7) (A8) (A9) (A10)**
6. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. O caso de uso é finalizado.

A2: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A3: Link Profile clicado.

1. Sistema chama o Use Case **UC010 – Alterar Conta do Usuário.**
2. O caso de uso é finalizado.

A4: Link Logout clicado.

1. Sistema realiza logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário.**
3. O caso de uso é finalizado.

A5: Link Jobs Overview clicado

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A6: Link Jobs Details clicado

1. Sistema chama o Use Case **UC005 – Visualizar Informações da Anotação.**
2. O caso de uso é finalizado.

A7: Link Distribution by Product clicado.

1. Sistema chama o Use Case **UC007 – Visualizar Agrupamentos por Produto.**
2. O caso de uso é finalizado.

A8: Link Distribution by Organism clicado.

1. Sistema chama o Use Case **UC008 – Visualizar Agrupamentos por Organism.**
2. O caso de uso é finalizado.

A9: Link COG Group clicado na tabela.

1. Sistema chama o **(A11).**

A10: Link COG Group clicado no gráfico.

1. Sistema chama o **(A11).**

A11: Fatia do gráfico ou link da tabela clicado.

1. Sistema seta novos parâmetros url.
2. O caso de uso é reiniciado.

Regras de Negócio

R1: O sistema realiza uma busca das informações agrupadas por COG Group e baseado nos parâmetros passados via url.

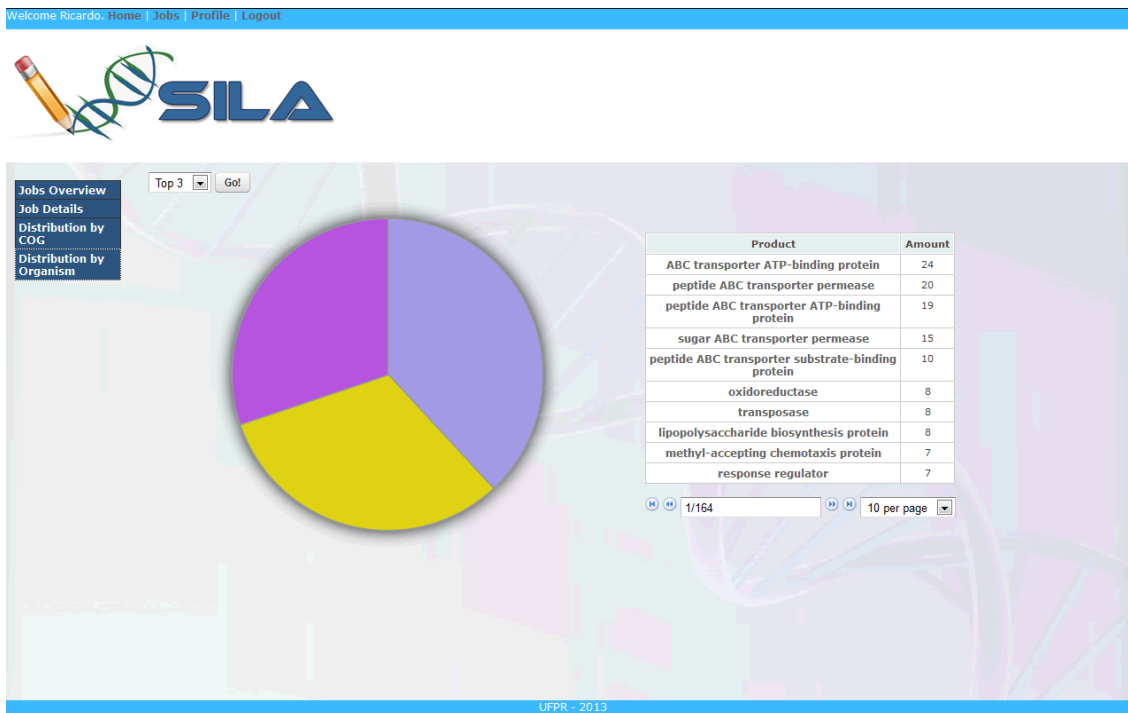
7.7 UC007 – Visualizar Agrupamentos por Produto

Descrição

Este caso de uso serve para o usuário visualizar os agrupamentos por Produto.

Data View

DV1 – Tela de visualização do gráfico de agrupamento por Produto.



Pré-condições

O usuário deverá estar logado no sistema e ter clicado no link “Distribution by Product”.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir visualizar os agrupamentos por Produto.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema faz a busca dos dados de agrupamentos por Produto, baseado nos parâmetros passados via url. **(R1)**
2. O sistema monta o gráfico com base nos dados retornados.
3. O sistema monta as tabelas com base nos dados retornados.
4. O sistema apresenta a tela **(DV1)**
5. O usuário visualiza as informações desejadas **(A1) (A2) (A3) (A4) (A5) (A6) (A7) (A8) (A9) (A10)**
6. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. O caso de uso é finalizado.

A2: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A3: Link Profile clicado.

1. Sistema chama o Use Case **UC010 – Alterar Conta do Usuário.**
2. O caso de uso é finalizado.

A4: Link Logout clicado.

1. Sistema realiza logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário.**
3. O caso de uso é finalizado.

A5: Link Jobs Overview clicado

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A6: Link Jobs Details clicado

1. Sistema chama o Use Case **UC005 – Visualizar Informações da Anotação.**
2. O caso de uso é finalizado.

A7: Link Distribution by COG clicado.

1. Sistema chama o Use Case **UC006 – Visualizar Agrupamentos por COG.**
2. O caso de uso é finalizado.

A8: Link Distribution by Organism clicado.

1. Sistema chama o Use Case **UC008 – Visualizar Agrupamentos por Organismo.**
2. O caso de uso é finalizado.

A9: Link Produto clicado na tabela.

1. Sistema chama o Use Case **UC009 – Visualizar Informação por Gene.**
2. O caso de uso é reiniciado.

A10: Link Produto clicado no gráfico.

1. Sistema chama o Use Case **UC009 – Visualizar Informação por Gene.**
2. O caso de uso é reiniciado.

Regras de Negócio

R1: O sistema realiza uma busca das informações agrupadas por Produto e baseado nos parâmetros passados via url.

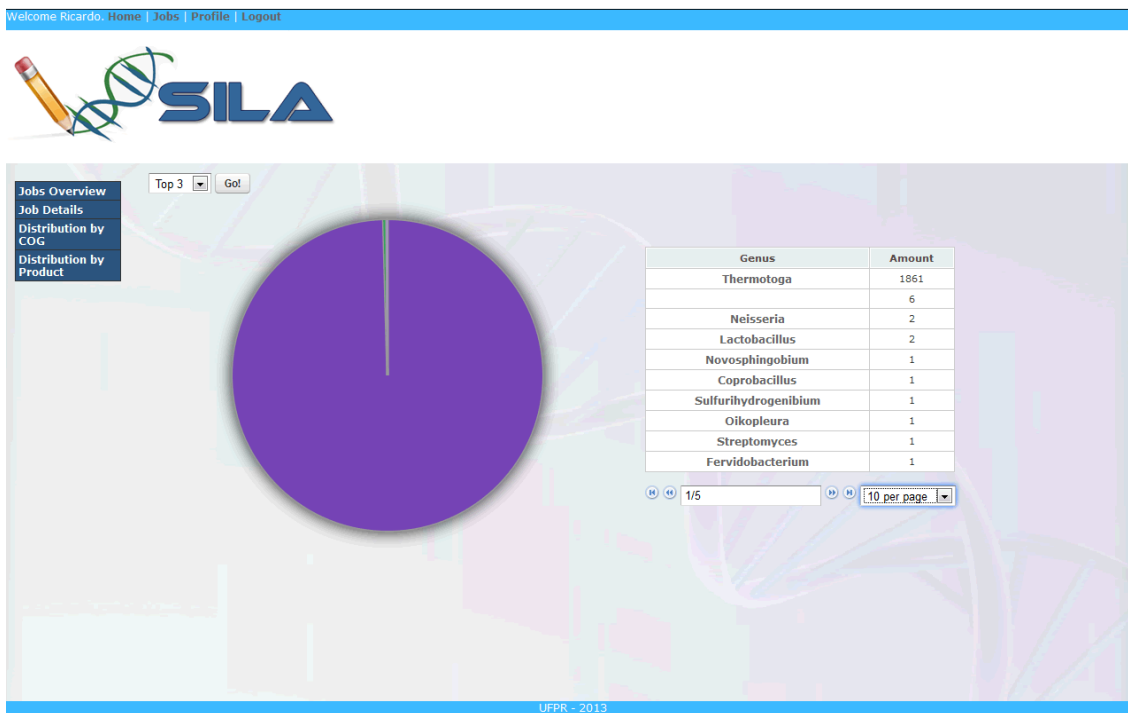
7.8 UC008 – Visualizar Agrupamentos por Organismo

Descrição

Este caso de uso serve para o usuário visualizar os agrupamentos por Organismo.

Data View

DV1 – Tela de visualização do gráfico de agrupamento por Organismo.



Pré-condições

O usuário deverá estar logado no sistema e ter clicado no link “Distribution by Organism”.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir visualizar os agrupamentos por Organismo.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema faz a busca dos dados de agrupamentos por Organismo, baseado nos parâmetros passados via url. **(R1)**
2. O sistema monta o gráfico com base nos dados retornados.
3. O sistema monta as tabelas com base nos dados retornados.
4. O sistema apresenta a tela **(DV1)**
5. O usuário visualiza as informações desejadas **(A1) (A2) (A3) (A4) (A5) (A6) (A7) (A8) (A9) (A10)**
6. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. O caso de uso é finalizado.

A2: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A3: Link Profile clicado.

1. Sistema chama o Use Case **UC010 – Alterar Conta do Usuário.**
2. O caso de uso é finalizado.

A4: Link Logout clicado.

1. Sistema realiza logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário.**
3. O caso de uso é finalizado.

A5: Link Jobs Overview clicado

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A6: Link Jobs Details clicado

1. Sistema chama o Use Case **UC005 – Visualizar Informações da Anotação.**
2. O caso de uso é finalizado.

A7: Link Distribution by Product clicado.

1. Sistema chama o Use Case **UC007 – Visualizar Agrupamentos por Produto.**
2. O caso de uso é finalizado.

A8: Link Distribution by COG clicado.

1. Sistema chama o Use Case **UC008 – Visualizar Agrupamentos por COG.**
2. O caso de uso é finalizado.

A9: Link Organismo clicado na tabela.

1. Sistema chama o **(A11).**

A10: Link Organismo clicado no gráfico.

1. Sistema chama o **(A11).**

A11: Fatia do gráfico ou link da tabela clicado.

1. Sistema seta novos parâmetros url.
2. O caso de uso é reiniciado.

Regras de Negócio

R1: O sistema realiza uma busca das informações agrupadas por Organismo e baseado nos parâmetros passados via url.

Usuário

Fluxo de Eventos Principal

1. O sistema faz a busca dos dados relacionados ao gene. **(R1)**
2. O sistema monta a tabela com base nos dados retornados.
3. O sistema apresenta a tela **(DV1)**
4. O usuário visualiza as informações desejadas **(A1) (A2) (A3) (A4) (A5) (A6) (A7) (A8) (A9) (A10)**
5. O caso de uso é finalizado.

Fluxos Alternativos

A1: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. O caso de uso é finalizado.

A2: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A3: Link Profile clicado.

1. Sistema chama o Use Case **UC010 – Alterar Conta do Usuário.**
2. O caso de uso é finalizado.

A4: Link Logout clicado.

1. Sistema realiza logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário.**
3. O caso de uso é finalizado.

A5: Link Jobs Overview clicado

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. O caso de uso é finalizado.

A6: Link Jobs Details clicado

1. Sistema chama o Use Case **UC005 – Visualizar Informações da Anotação.**
2. O caso de uso é finalizado.

A7: Link Distribution by Product clicado.

1. Sistema chama o Use Case **UC007 – Visualizar Agrupamentos por Produto.**
2. O caso de uso é finalizado.

A8: Link Distribution by COG clicado.

1. Sistema chama o Use Case **UC006 – Visualizar Agrupamentos por COG.**
2. O caso de uso é finalizado.

A9: Link Distribution by Organism clicado.

1. Sistema chama o Use Case **UC008 – Visualizar Agrupamentos por Organismo.**
2. O caso de uso é finalizado.

A10: Link Enzima clicado.

1. Sistema apresenta a tela **(DV2).**

2. O Usuário visualiza as informações.
3. O caso de uso é encerrado.

Regras de Negócio

R1: O sistema realiza uma busca das informações relacionadas ao gene na base de dados e no arquivo .GBK.

7.10 UC010 – Alterar Conta do Usuário

Descrição

Este caso de uso serve para alterar os dados do usuário.

Data View

DV1 – Tela para alterar dados do usuário.

Pré-condições

O usuário deverá estar logado no sistema.

Pós-condições

Após o fim normal deste caso de uso:

1. O usuário deverá conseguir alterar os dados cadastrados.

Ator primário

Usuário

Fluxo de Eventos Principal

1. O sistema busca do banco de dados o nome, email e login do usuário que está logado.
2. O sistema apresenta a tela **(DV1)** preenchida com as informações obtidas no passo 1.
3. O usuário atualiza dados cadastrais. **(E1)(E2)(E3)(E4)(E5)(E6)(E7)**
4. O usuário pressiona o botão Submit **(A1)(A2)(A3)(A4)(A5)**
5. Sistema atualiza dados do usuário no banco de dados.
6. O sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
7. O caso de uso é finalizado.

Fluxos Alternativos

A1: Botão Back clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. Caso de uso é finalizado.

A2: Link Home clicado.

1. Sistema chama o Use Case **UC003 – Enviar Nova Tarefa.**
2. Caso de uso é finalizado.

A3: Link Jobs clicado.

1. Sistema chama o Use Case **UC004 – Visualizar Tarefas do Usuário.**
2. Caso de uso é finalizado.

A4: Link Profile clicado.

1. Caso de uso é reiniciado.

A5: Link Logout clicado.

1. Sistema efetua o logoff do usuário.
2. Sistema chama o Use Case **UC002 – Realizar Login de Usuário.**
3. Caso de uso é finalizado.

Fluxos de Exceção

E1: Campo Name não preenchido.

1. Sistema apresenta a mensagem “Please provide a name” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E2: Campo Email não preenchido.

1. Sistema apresenta a mensagem “Please provide a valid email address” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E3: Campo Email preenchido incorretamente.

1. Sistema consiste os dados inseridos no campo Email
2. Sistema apresenta a mensagem “Please provide a valid email address” ao usuário.
3. O usuário pressiona o botão OK da mensagem.
4. Retorna ao fluxo principal.

E4: Campo Login não preenchido.

1. Sistema apresenta a mensagem “Please provide a login” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E5: Campo Password não preenchido.

1. Sistema apresenta a mensagem “Please provide a password” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E6: Campo Retype password não preenchido.

1. Sistema apresenta a mensagem “Please retype your password” ao usuário.
2. O usuário pressiona o botão OK da mensagem.
3. Retorna ao fluxo principal.

E7: Campo Password e Retype password preenchidos com valores diferentes.

1. Sistema apresenta a mensagem “The provided passwords are different.” ao usuário.
2. O usuário pressiona o botão OK da mensagem.

3. Retorna ao fluxo principal.

Regras de Negócio

R1. O campo Email deve ser preenchido com um endereço de e-mail válido.

8 DIAGRAMAS DE SEQUÊNCIA

8.1 UC001 - Cadastrar Usuário

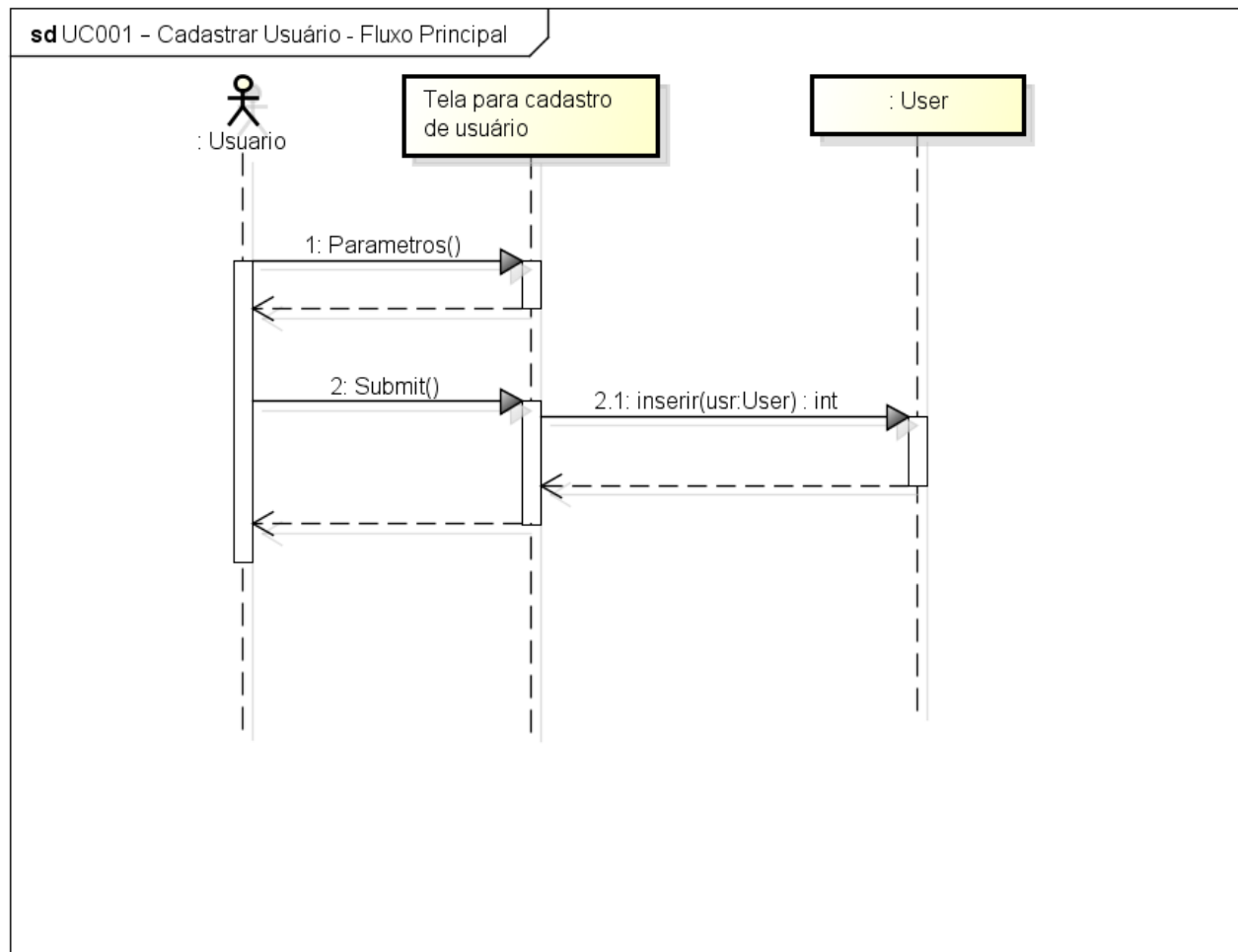


FIGURA 21 - DIAGRAMA DE SEQUÊNCIA - CADASTRAR USUÁRIO - FLUXO PRINCIPAL

FONTE: Os autores (2013)

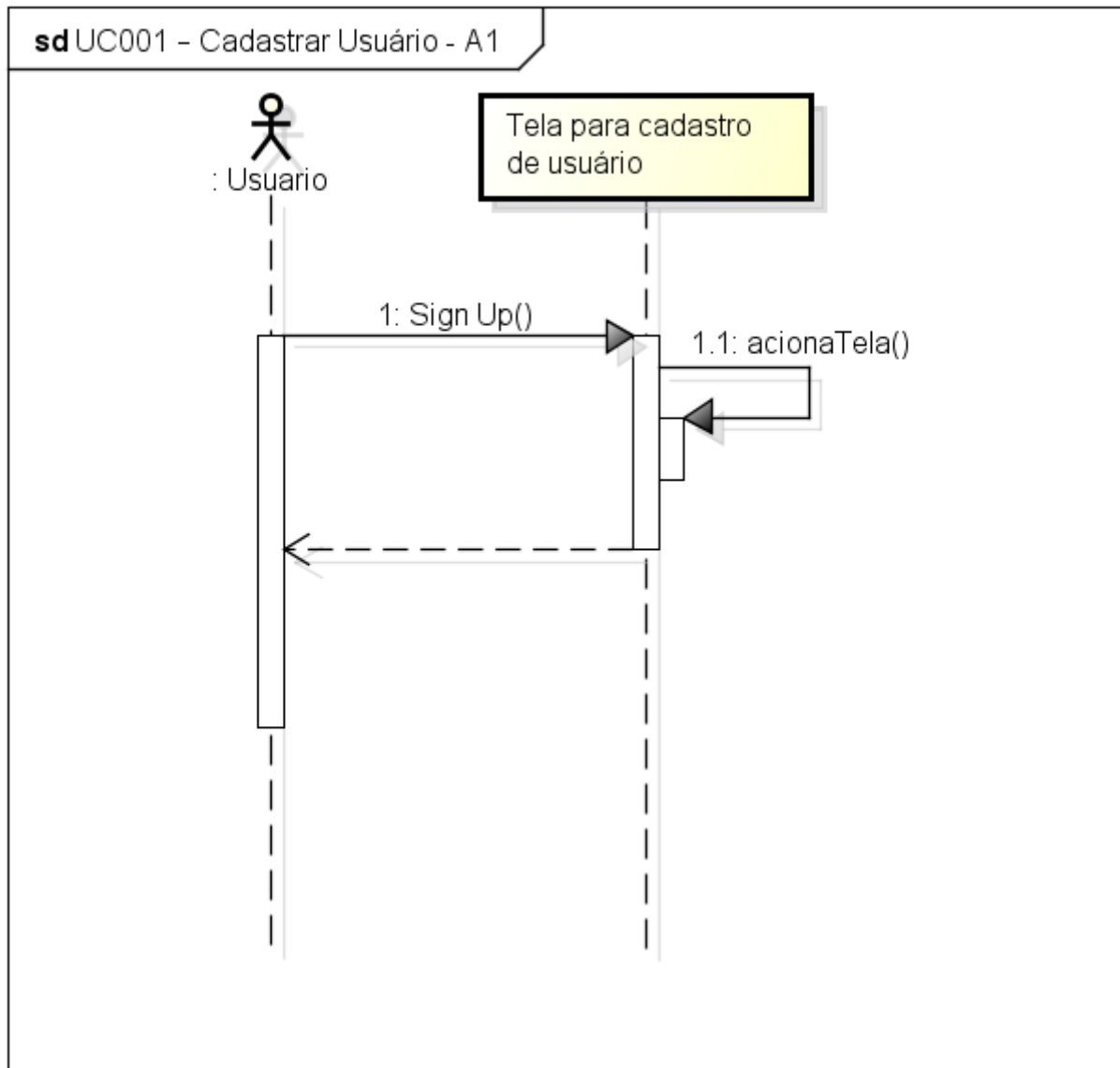


FIGURA 22 - DIAGRAMA DE SEQUÊNCIA - CADASTRAR USUÁRIO - FLUXO A1

FONTE: Os autores (2013)

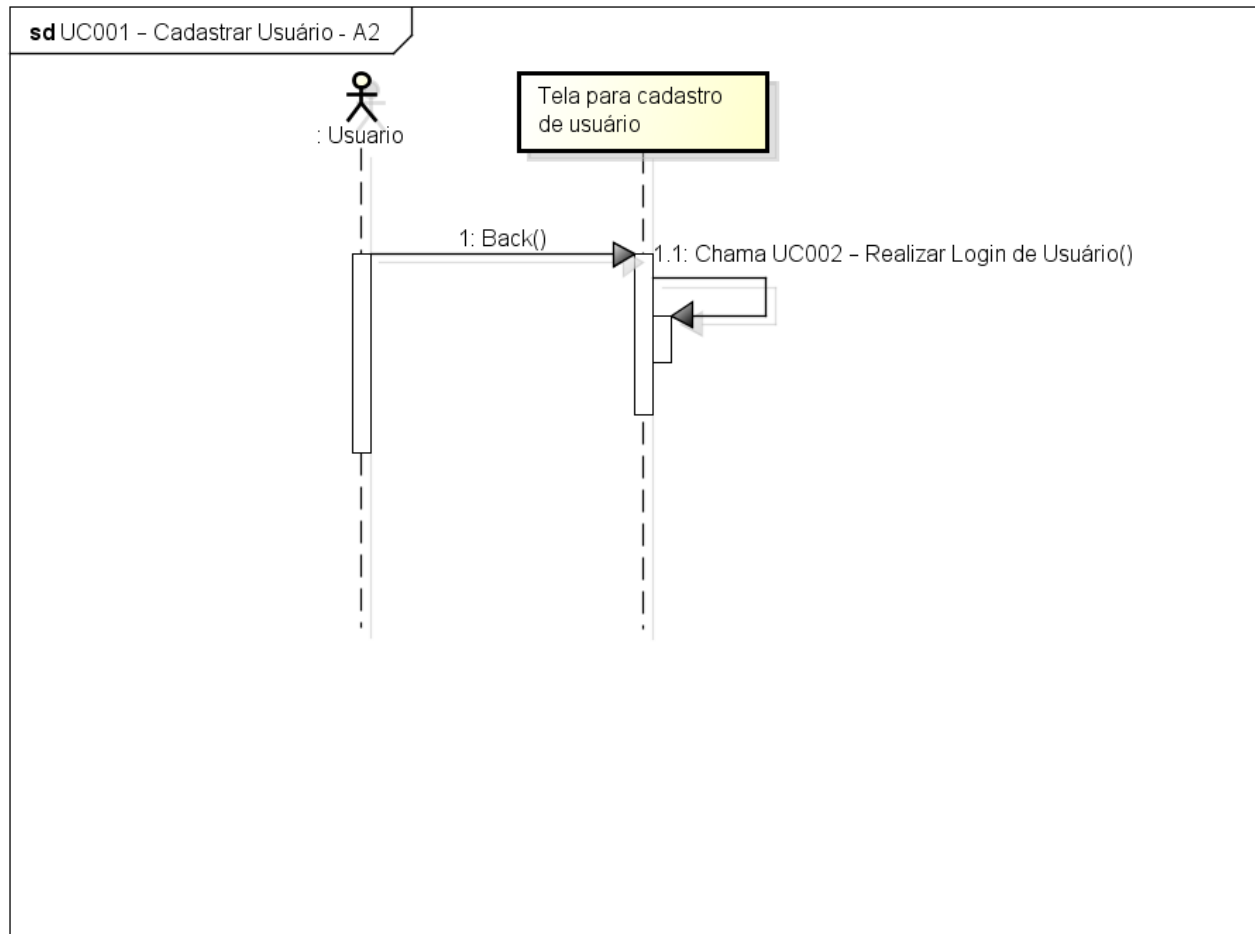


FIGURA 23 - DIAGRAMA DE SEQUÊNCIA - CADASTRAR USUÁRIO - FLUXO A2

FONTE: Os autores (2013)

8.2 UC002 - Realizar Login de Usuário

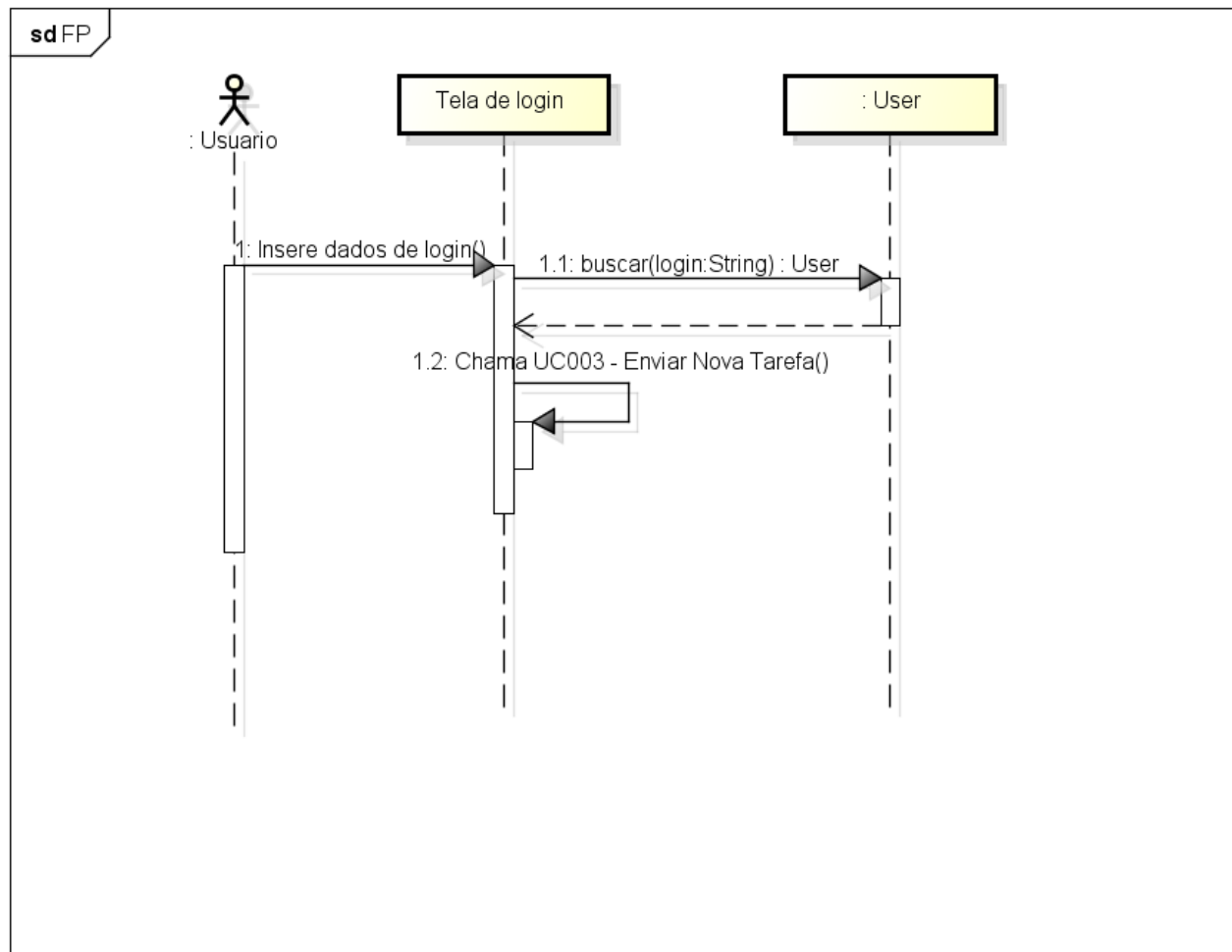


FIGURA 24 - DIAGRAMA DE SEQUENCIA - REALIZAR LOGIN DE USUÁRIO - FLUXO PRINCIPAL

FONTE: Os autores (2013)

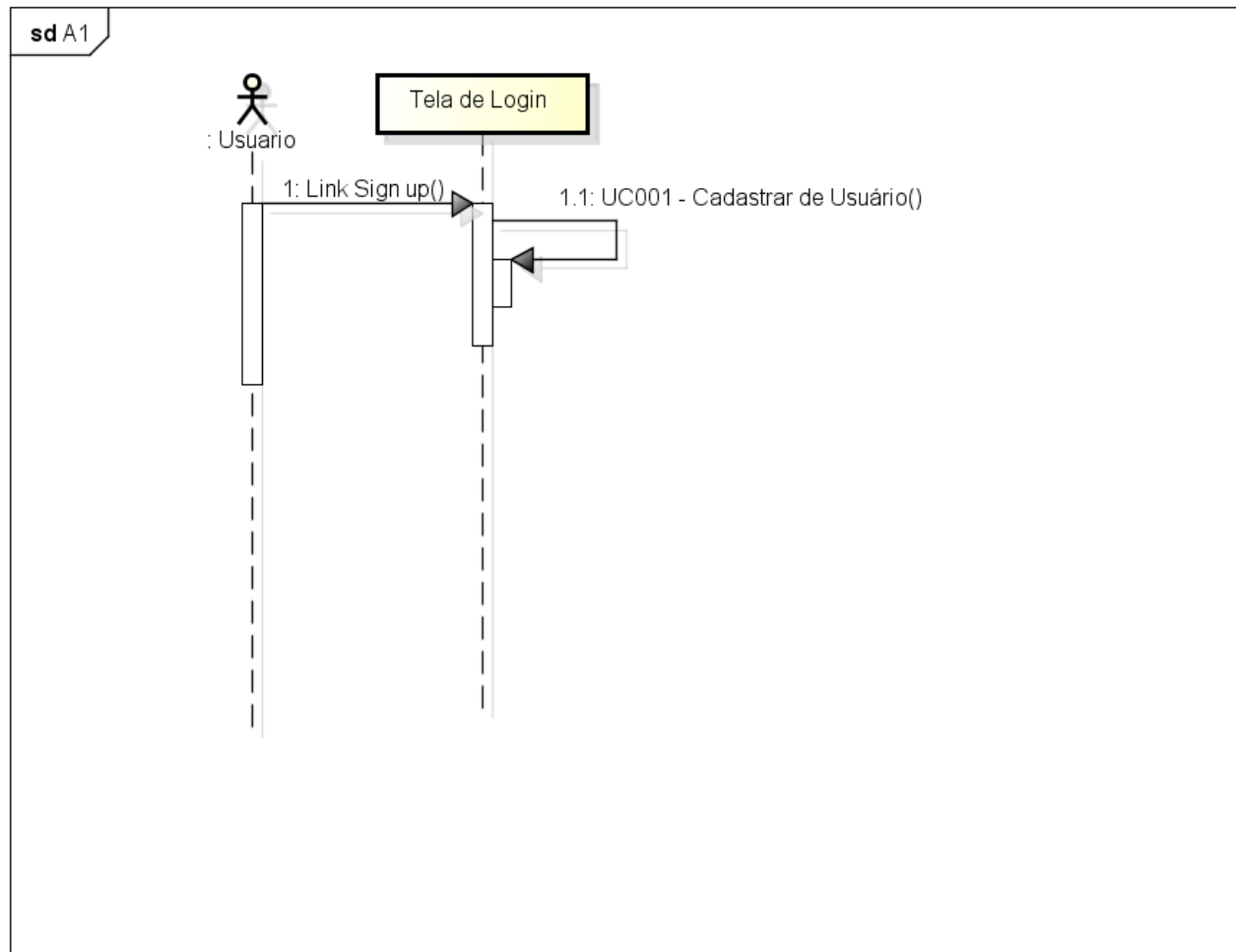


FIGURA 25 - DIAGRAMA DE SEQUENCIA - REALIZAR LOGIN DE USUÁRIO - FLUXO A1

FONTE: Os autores (2013)

8.3 UC003 - Enviar Nova Tarefa

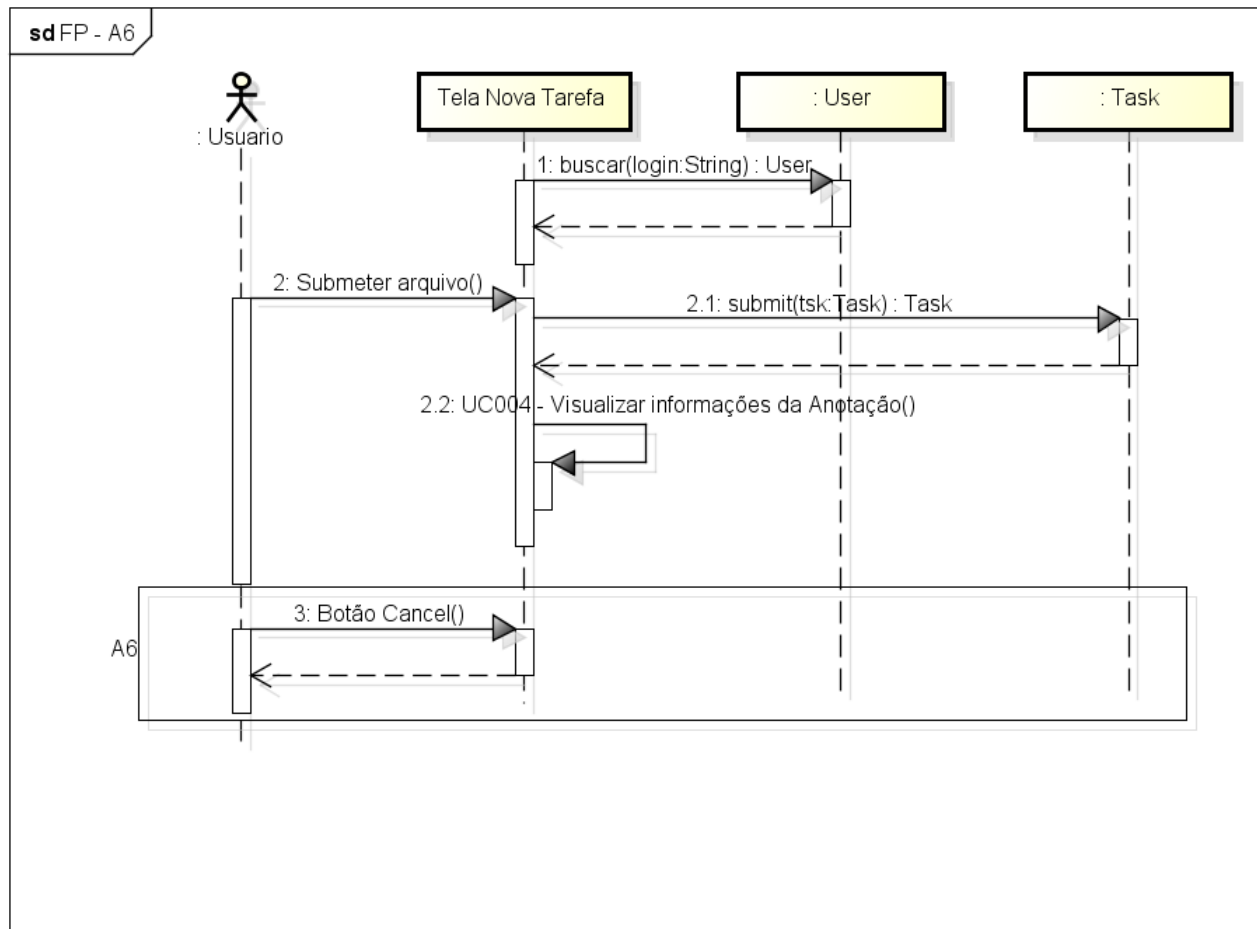


FIGURA 26 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO PRINCIPAL E A6

FONTE: Os autores (2013)

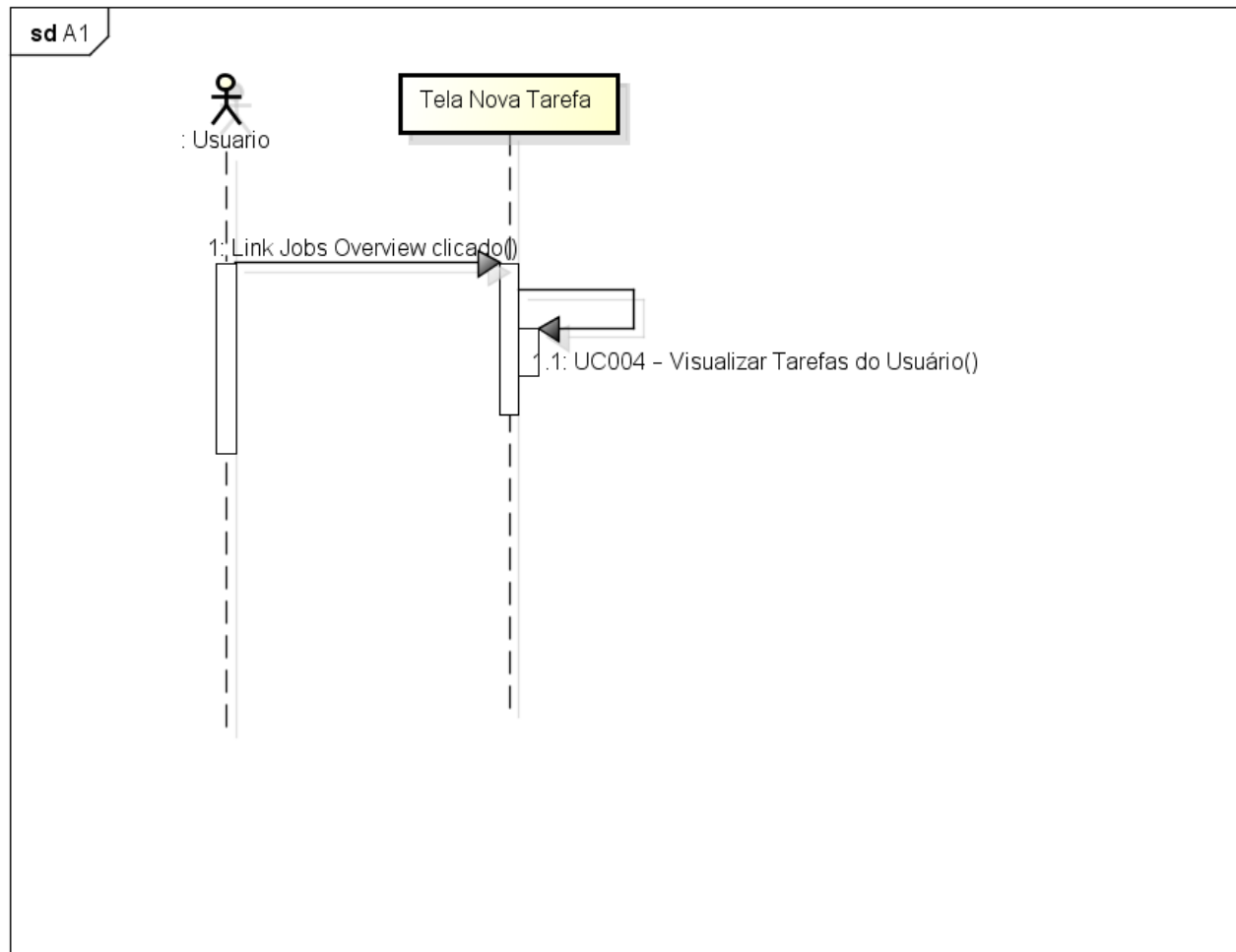


FIGURA 27 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A1

FONTE: Os autores (2013)

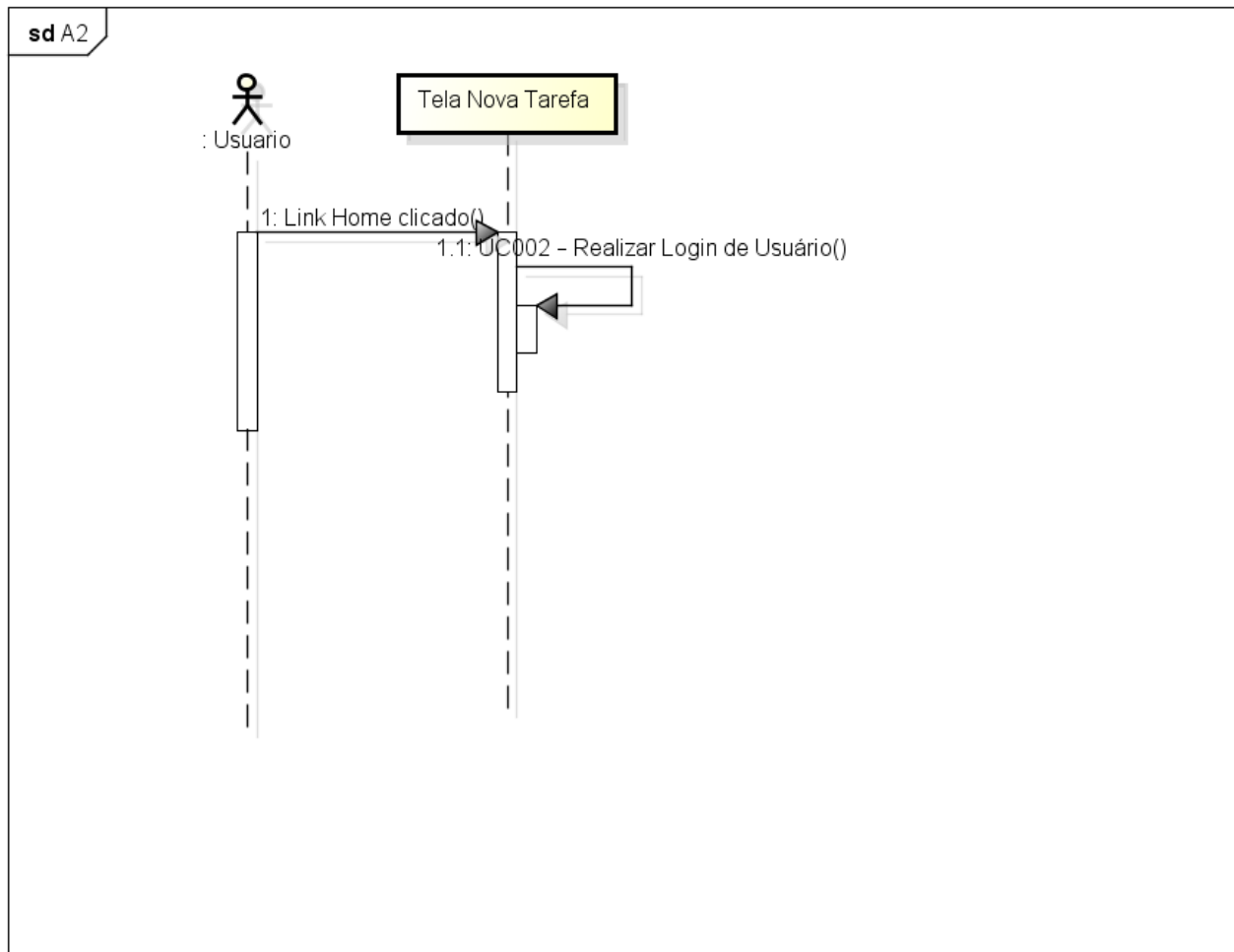


FIGURA 28 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A2

FONTE: Os autores (2013)

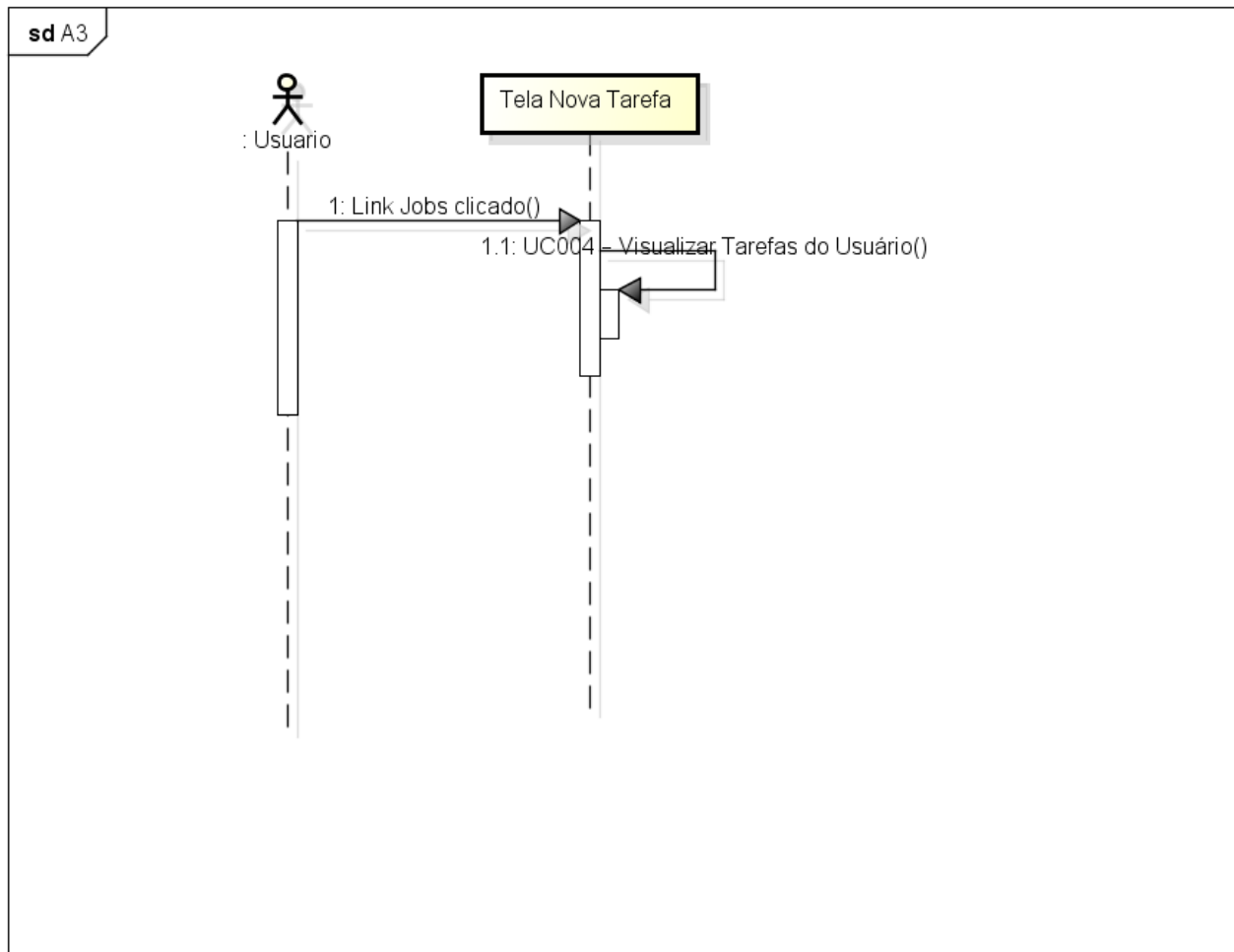


FIGURA 29 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A3

FONTE: Os autores (2013)

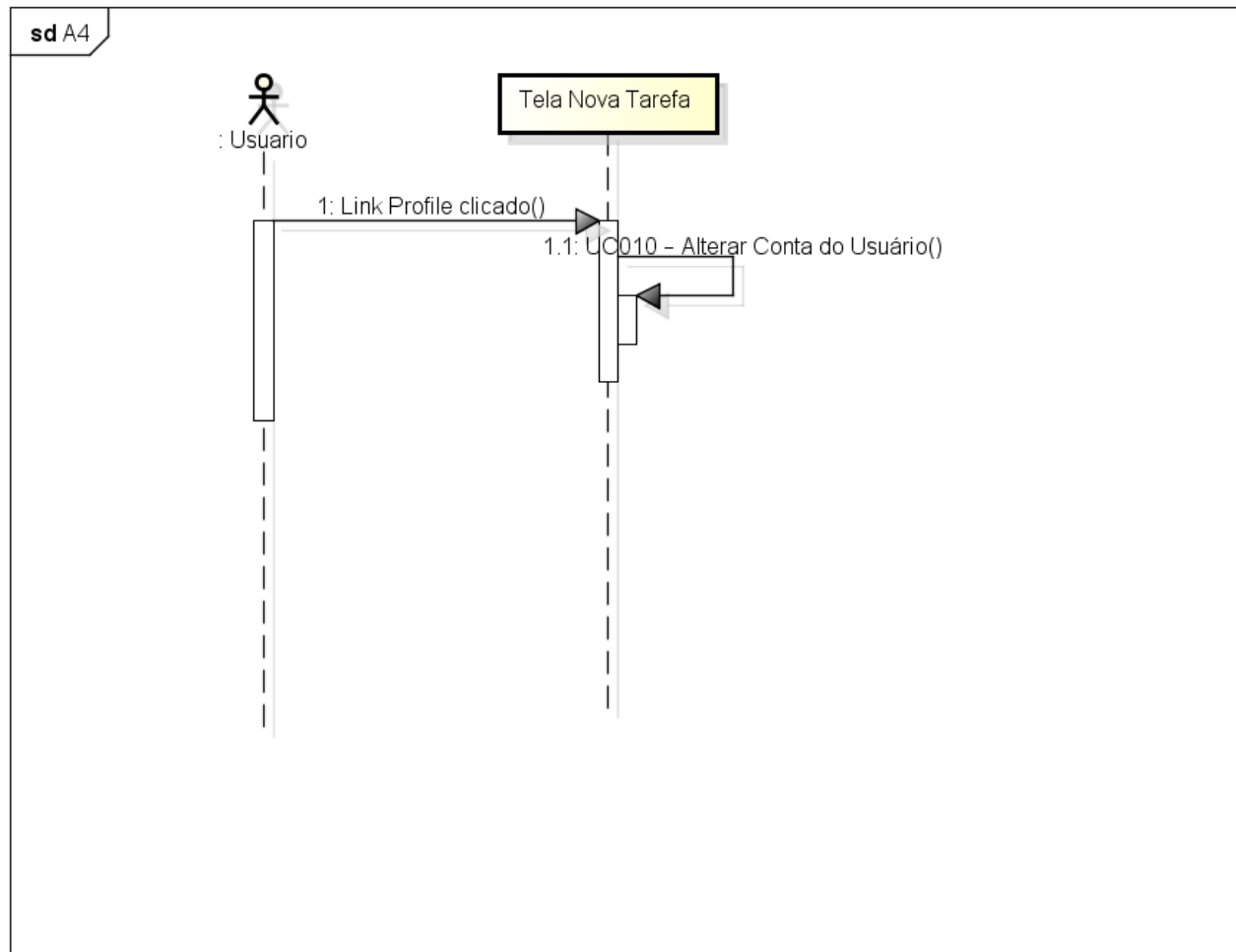


FIGURA 30 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A4

FONTE: Os autores (2013)

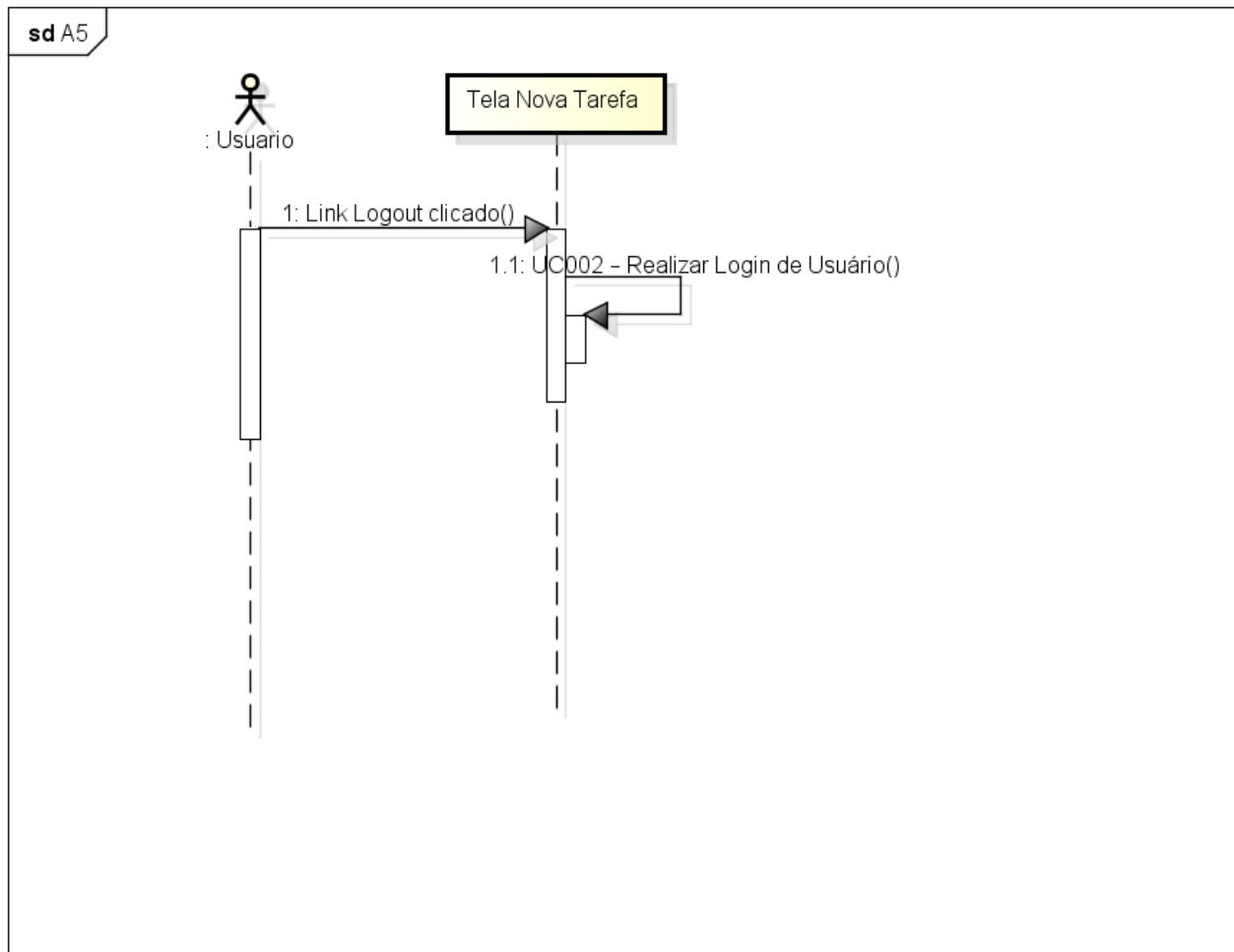


FIGURA 31 - DIAGRAMA DE SEQUENCIA - ENVIAR NOVA TAREFA - FLUXO A5

FONTE: Os autores (2013)

8.4 UC004 - Visualizar Tarefas do Usuário

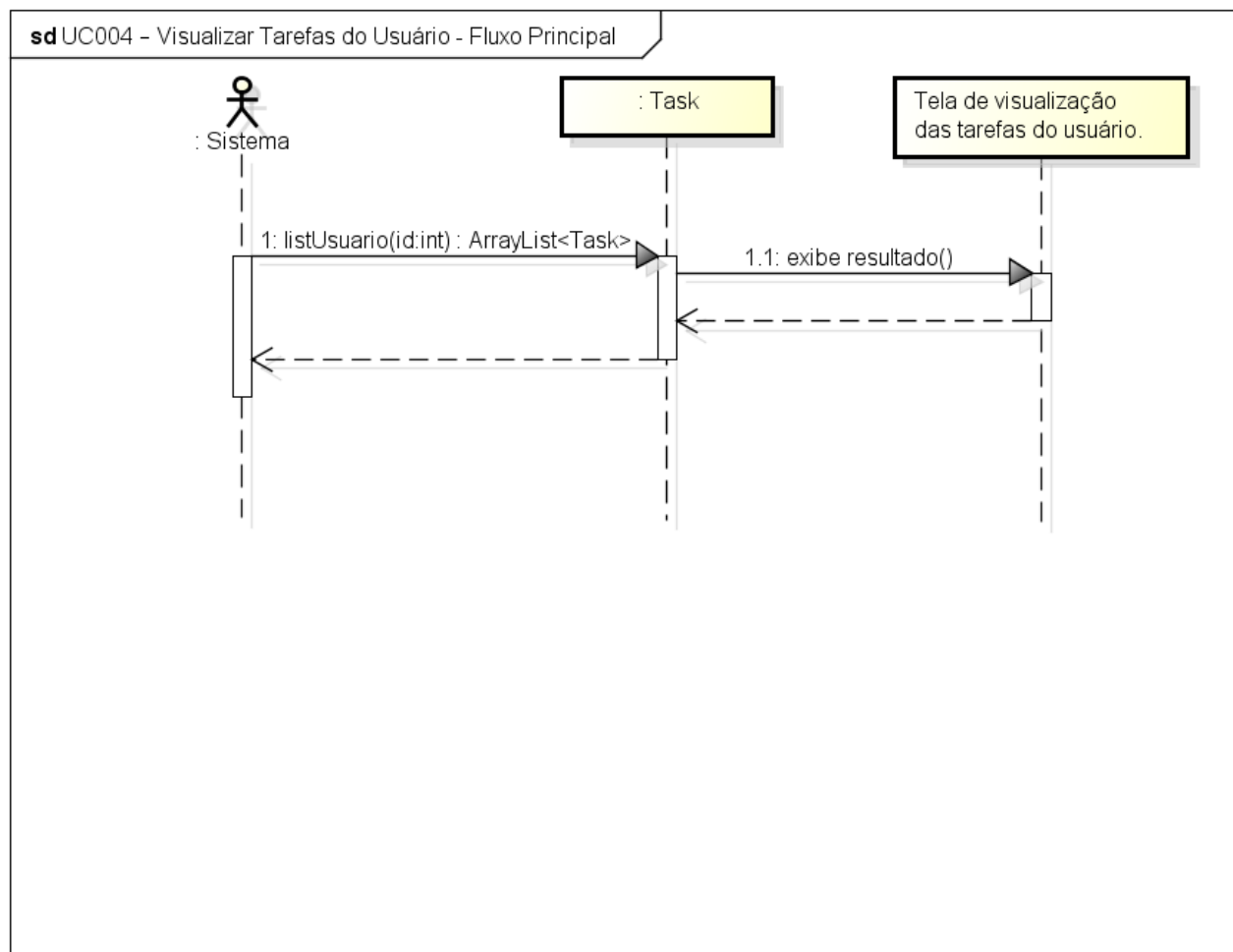


FIGURA 32 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO PRINCIPAL
FONTE: Os autores (2013)

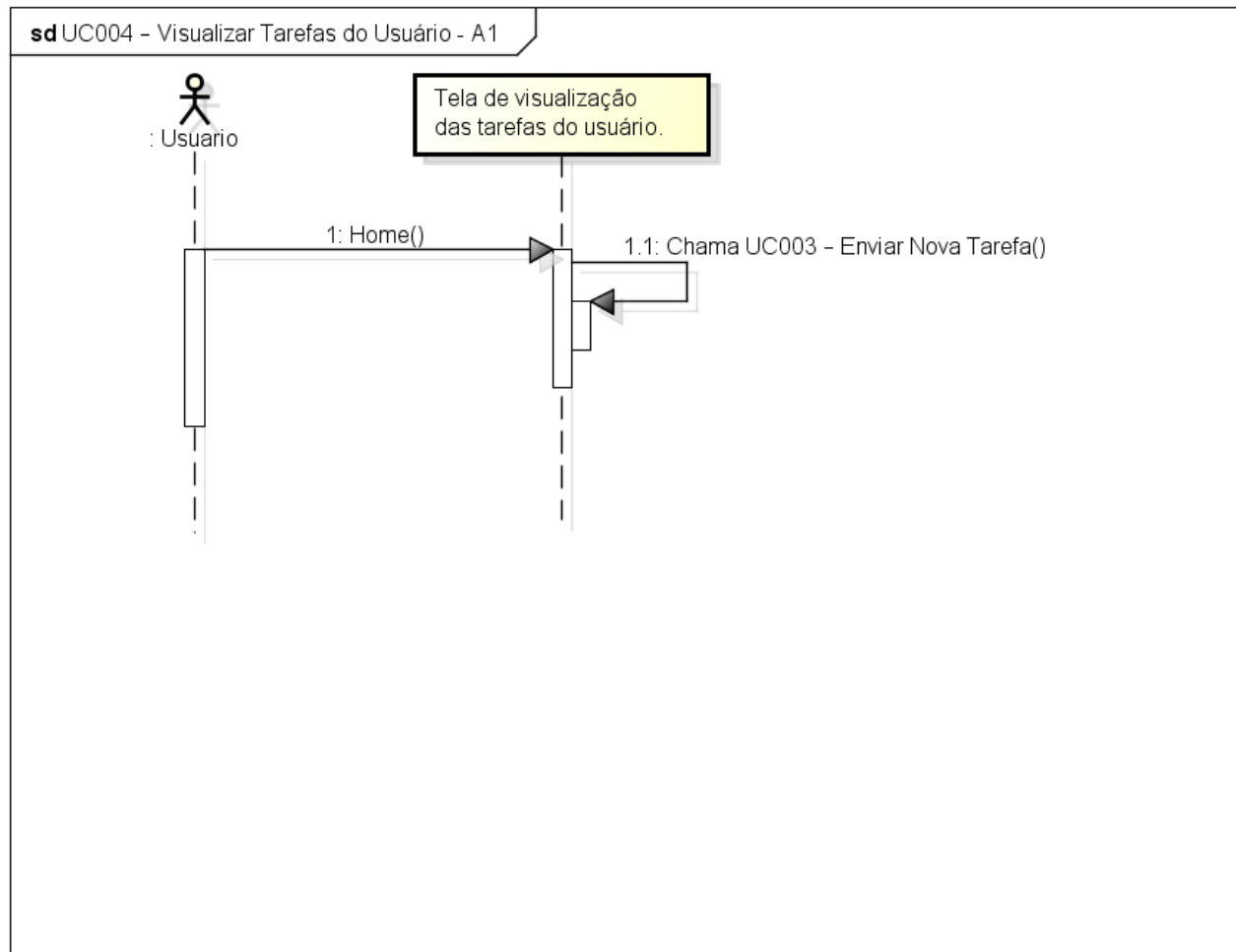


FIGURA 33 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A1

FONTE: Os autores (2013)

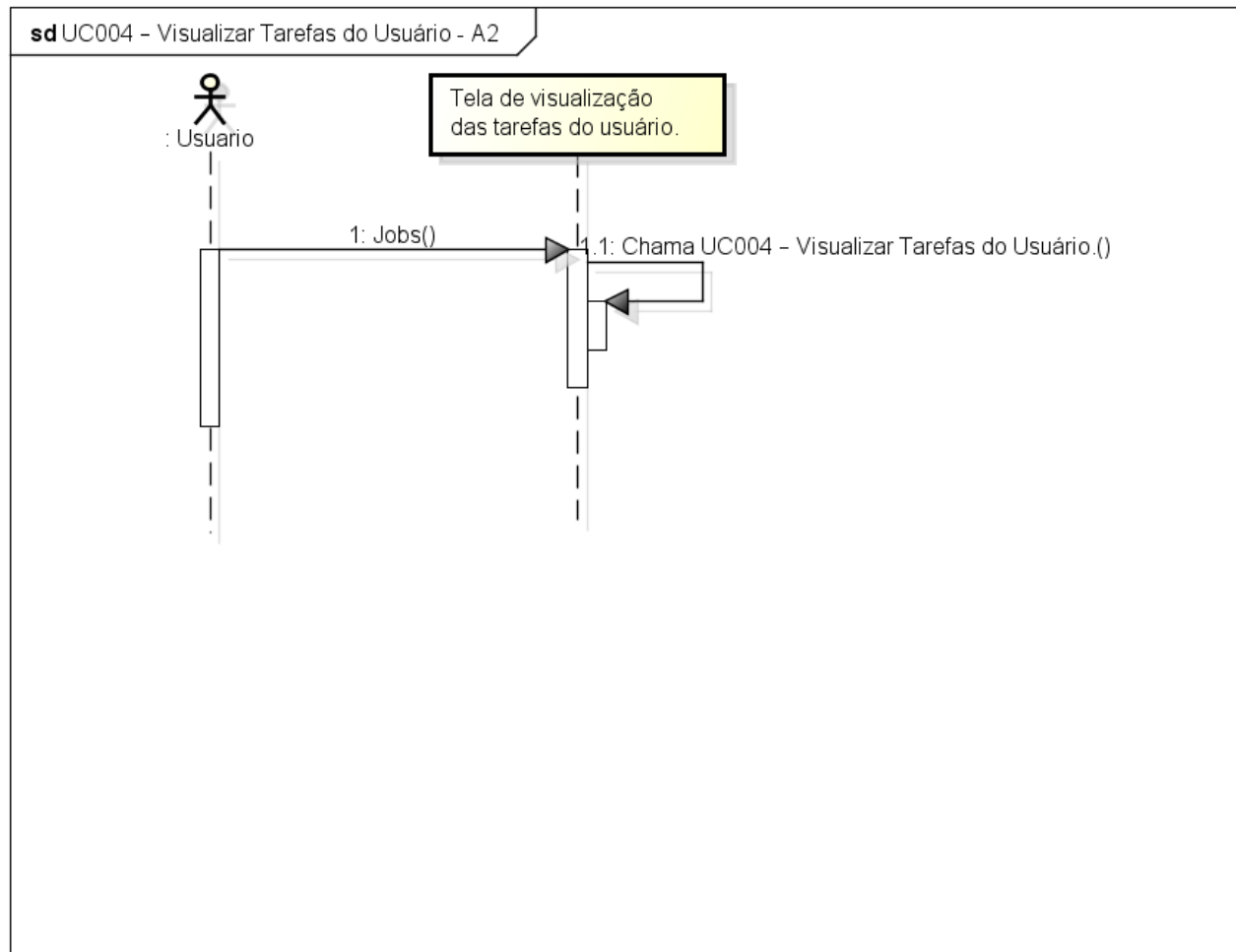


FIGURA 34 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A2

FONTE: Os autores (2013)

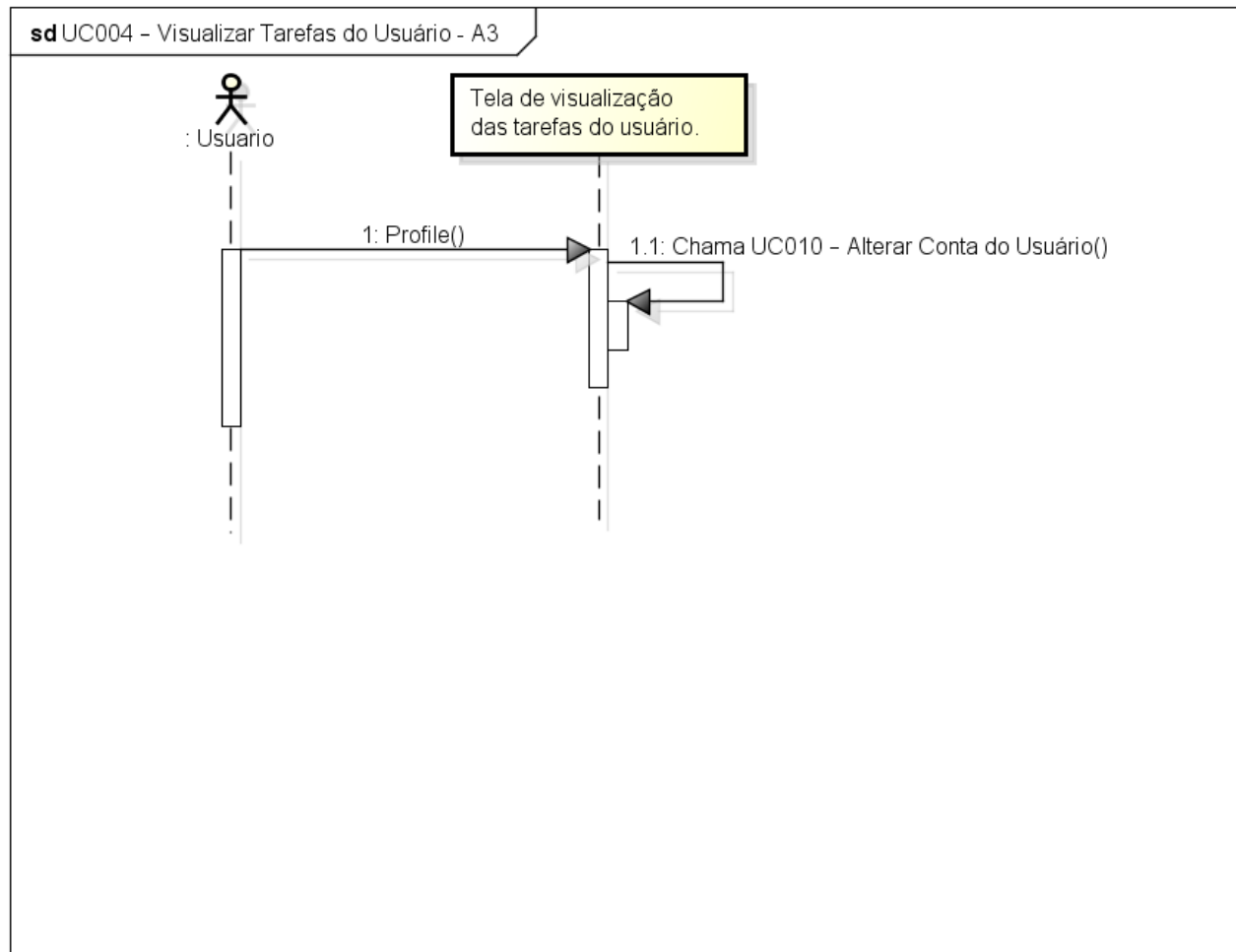


FIGURA 35 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A3

FONTE: Os autores (2013)

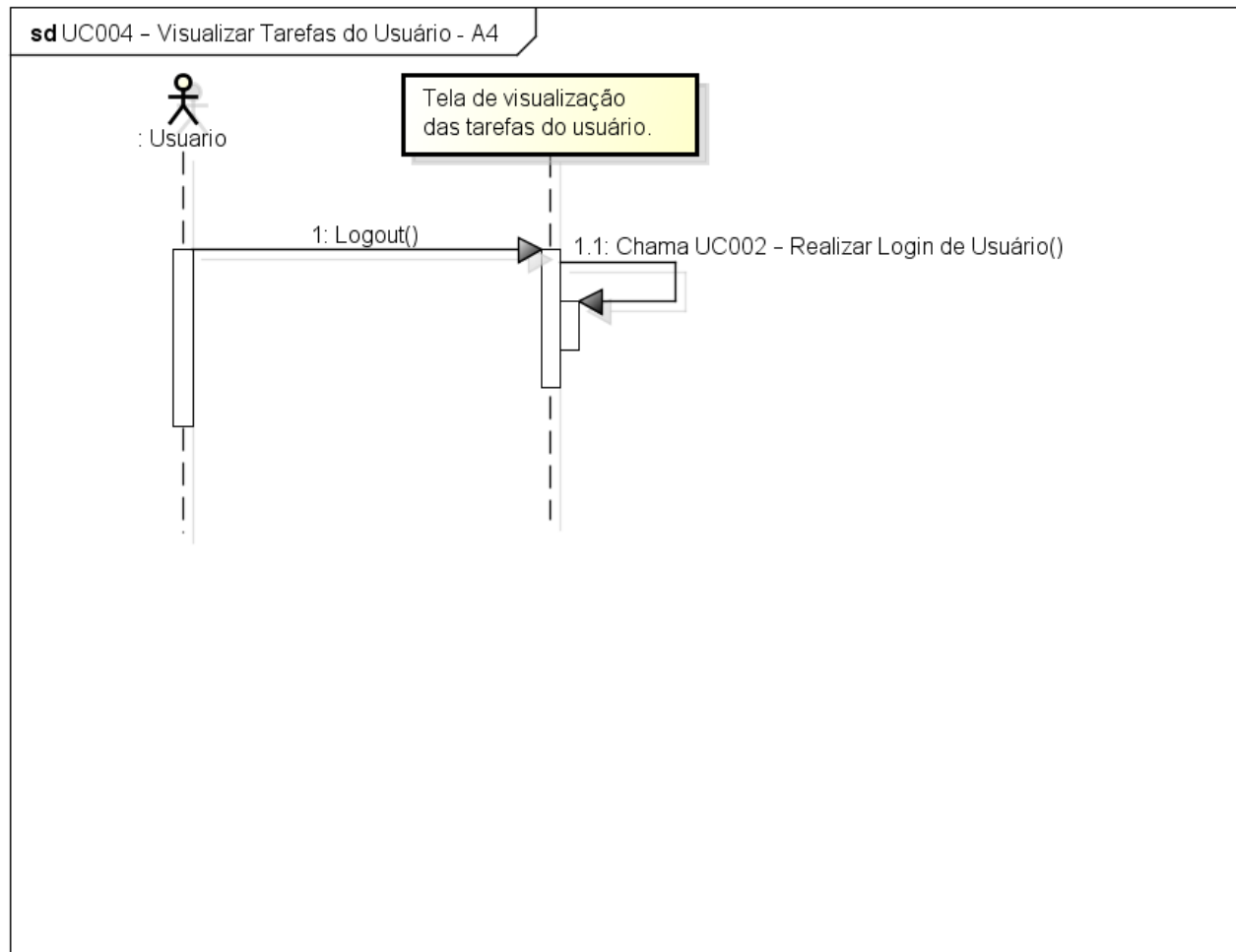


FIGURA 36 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A4

FONTE: Os autores (2013)

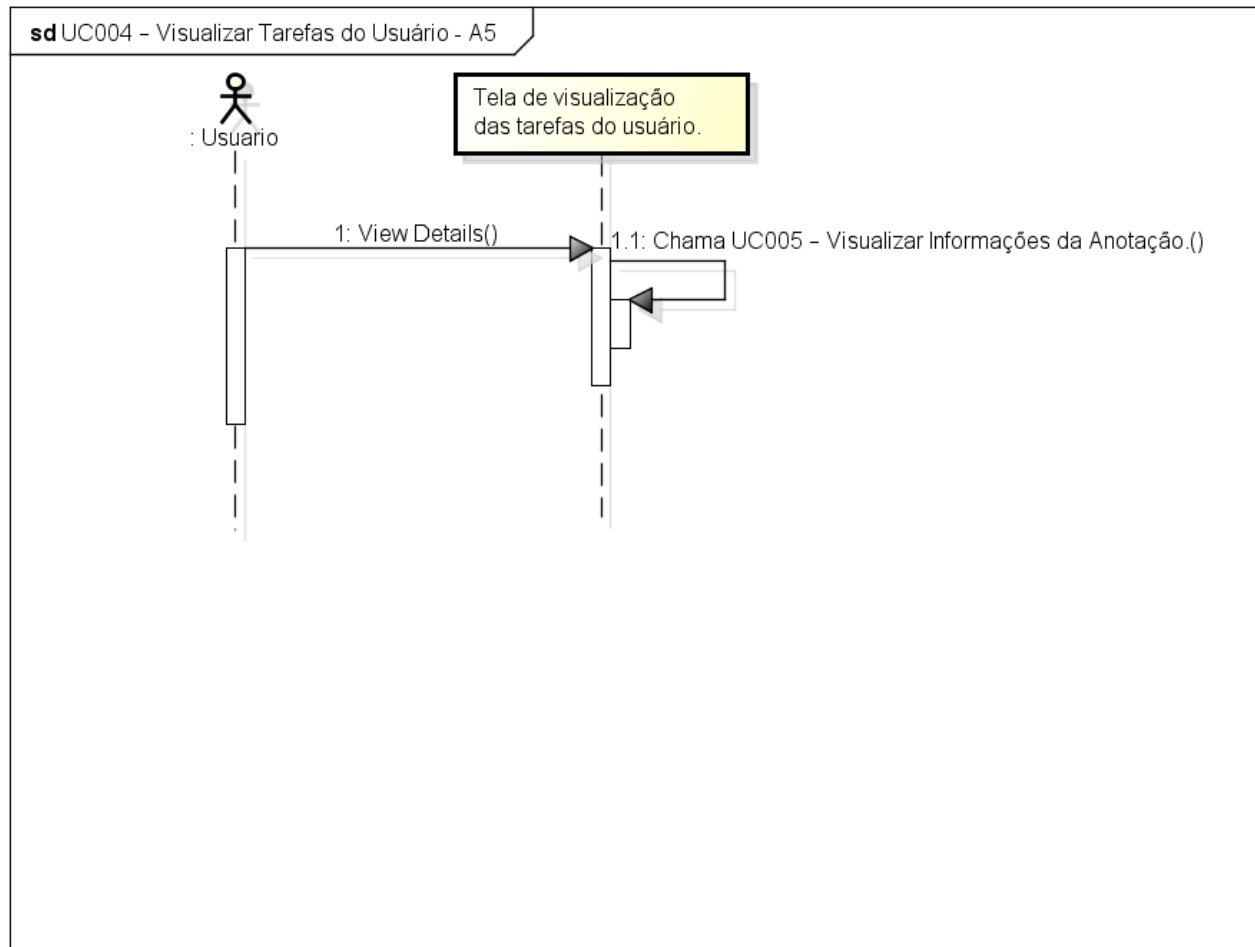


FIGURA 37 - DIAGRAMA DE SEQUENCIA - VISUALIZAR TAREFAS DO USUÁRIO - FLUXO A5

FONTE: Os autores (2013)

8.5 UC005 - Visualizar Informações da Anotação

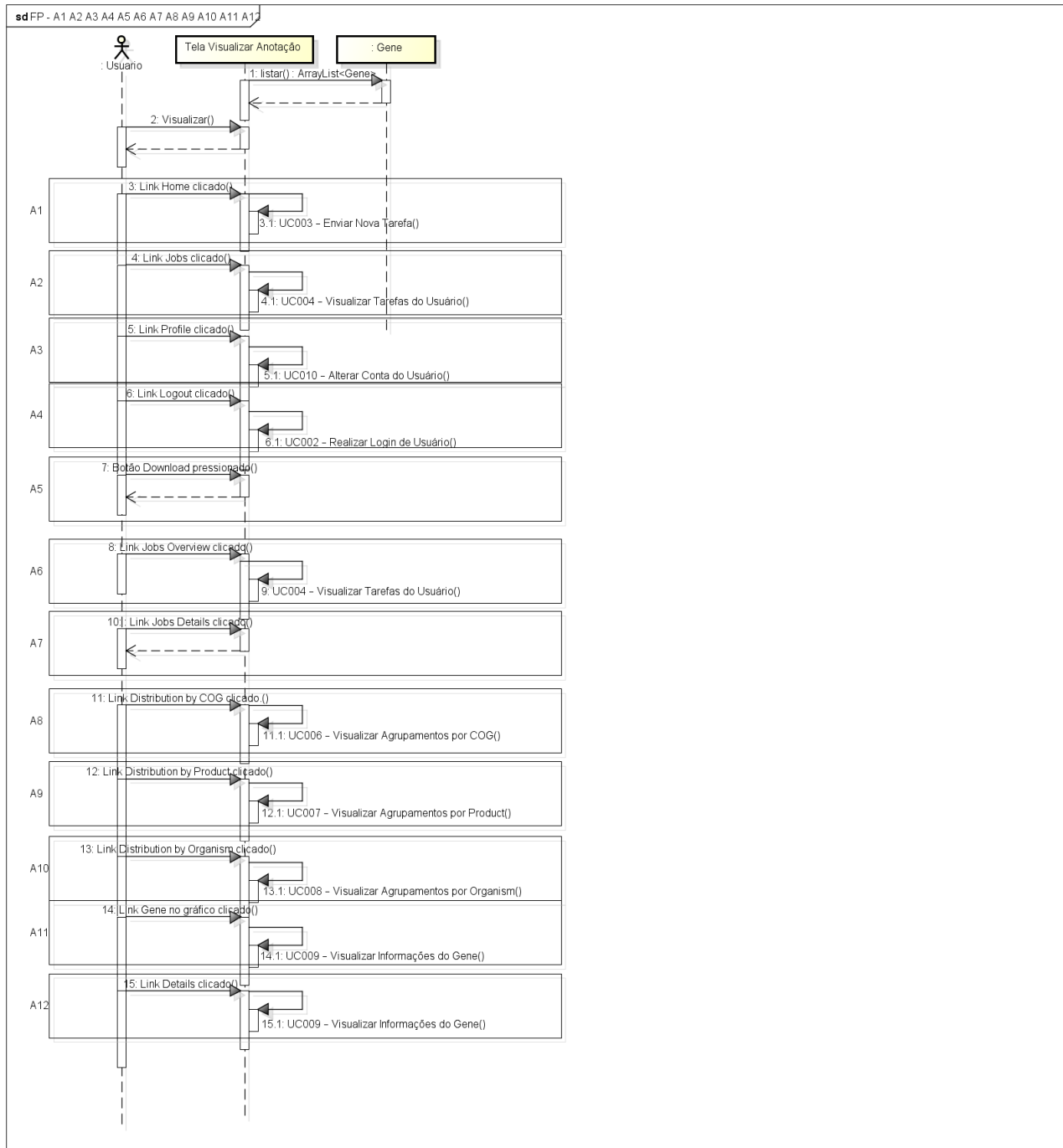


FIGURA 38 - DIAGRAMA DE SEQUENCIA - VISUALIZAR INFORMAÇÕES DA ANOTAÇÃO - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11 e A12

FONTE: Os autores (2013)

8.6 UC006 - Visualizar Agrupamentos por COG

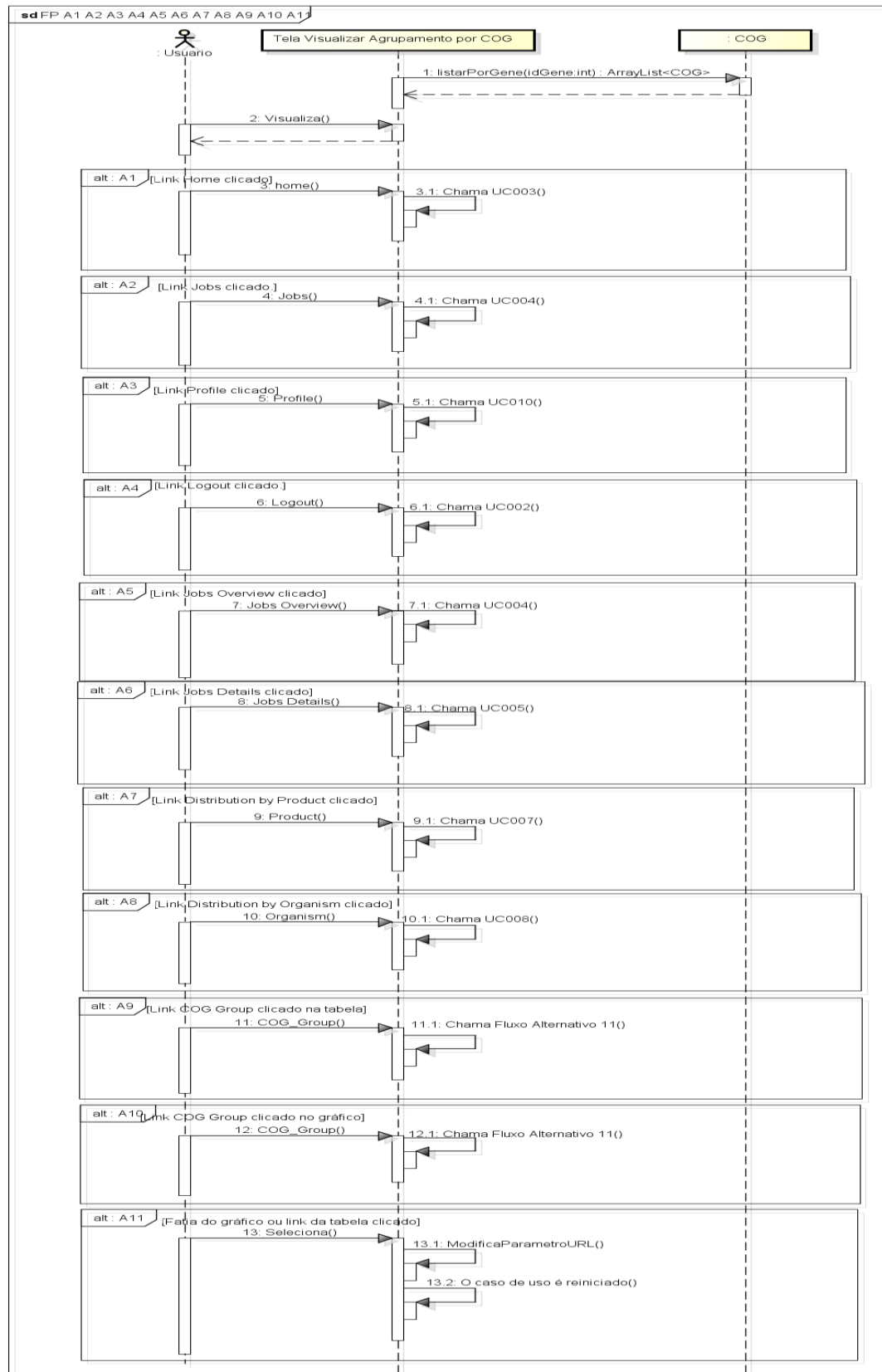


FIGURA 39 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTO POR COG - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10 e A11

FONTE: Os autores (2013)

8.7 UC007 - Visualizar Agrupamentos por Produto

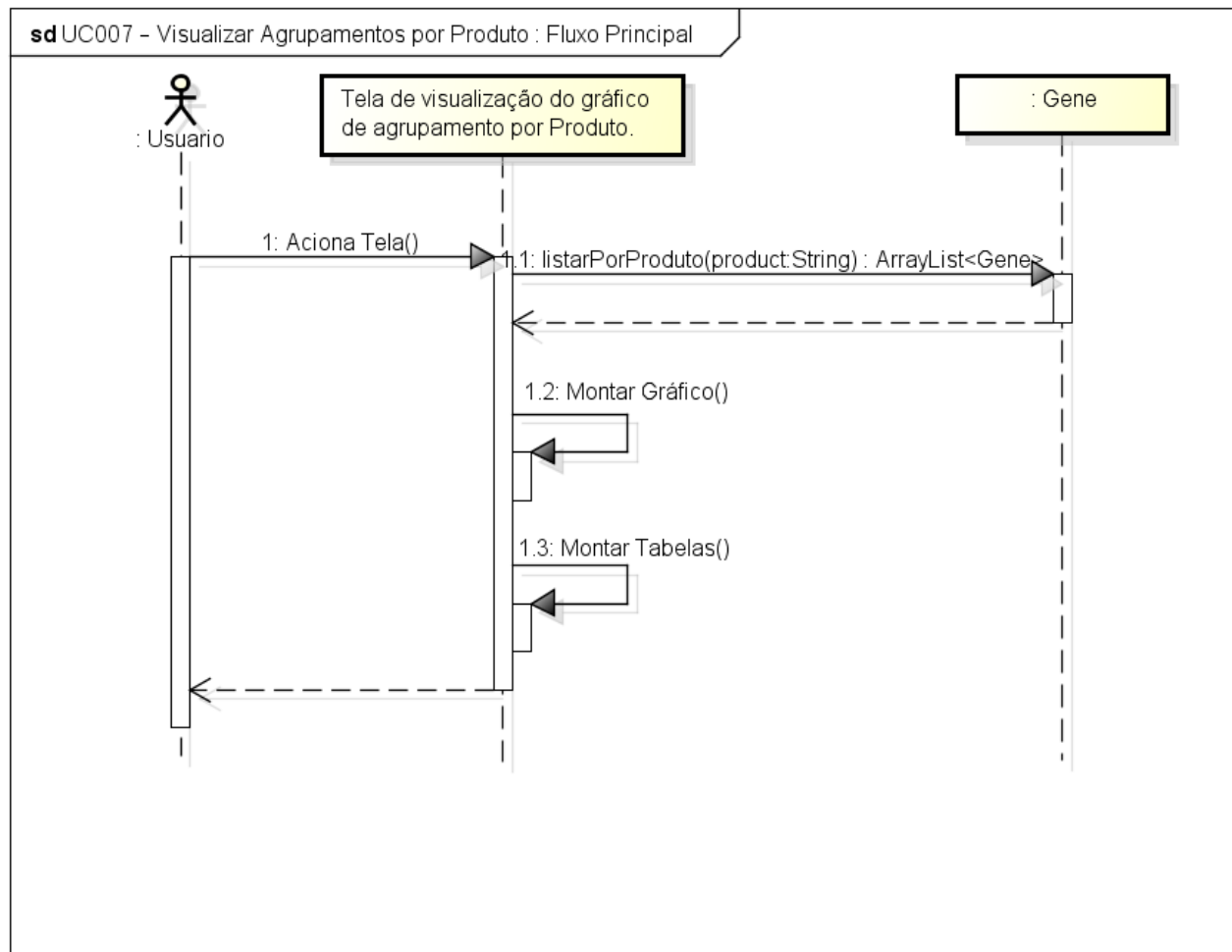


FIGURA 40 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO PRINCIPAL

FONTE: Os autores (2013)

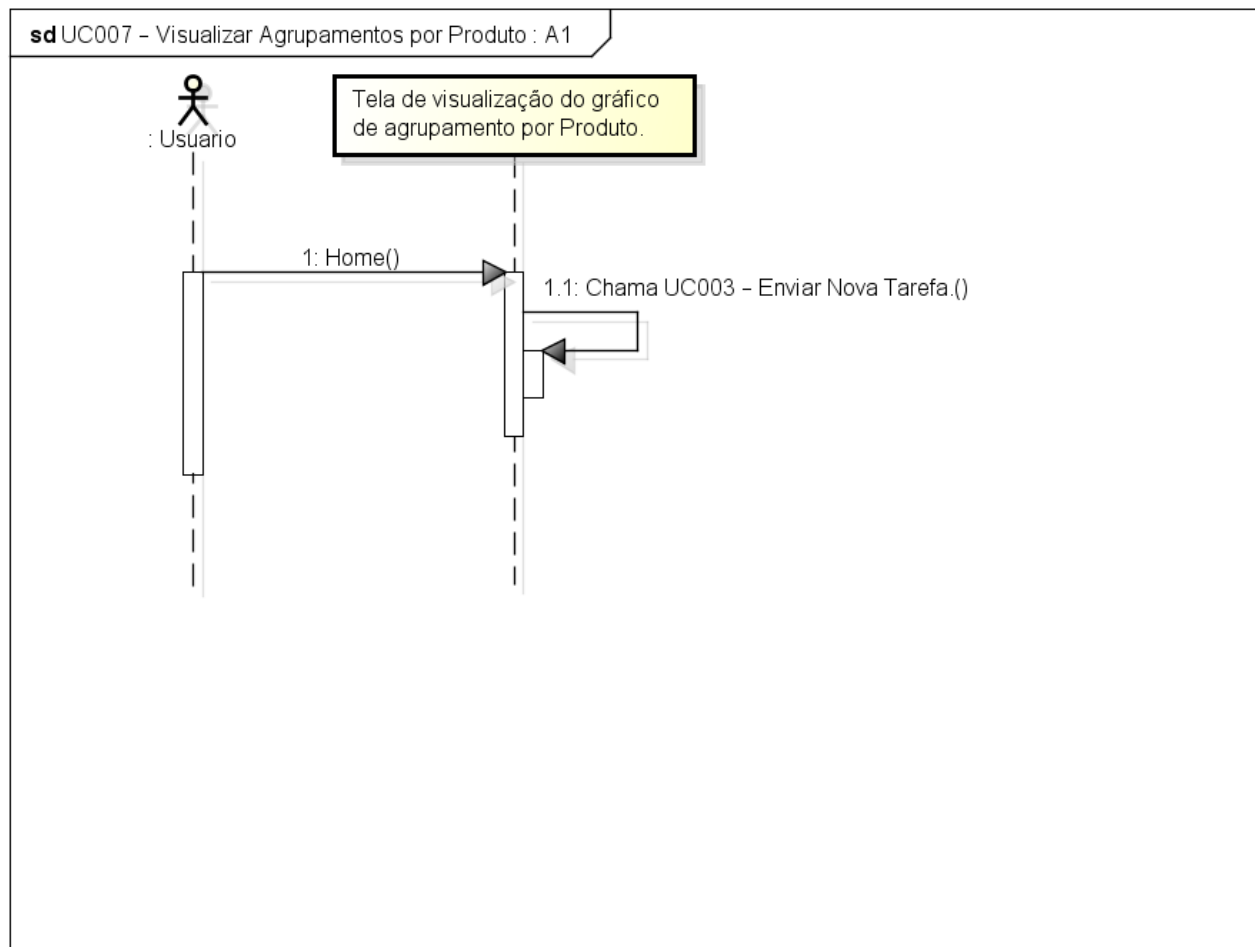


FIGURA 41 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A1
FONTE: Os autores (2013)

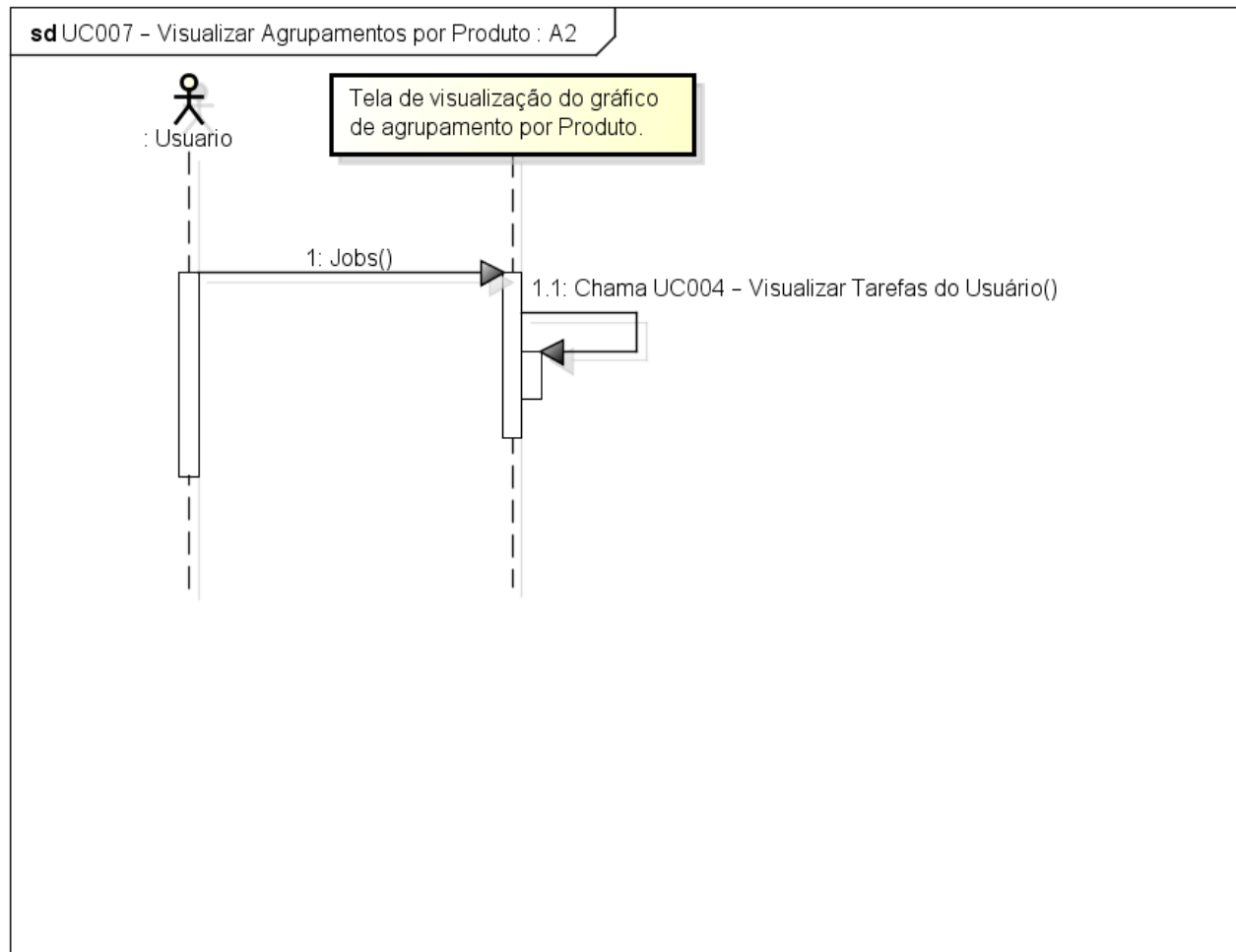


FIGURA 42 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A2
FONTE: Os autores (2013)

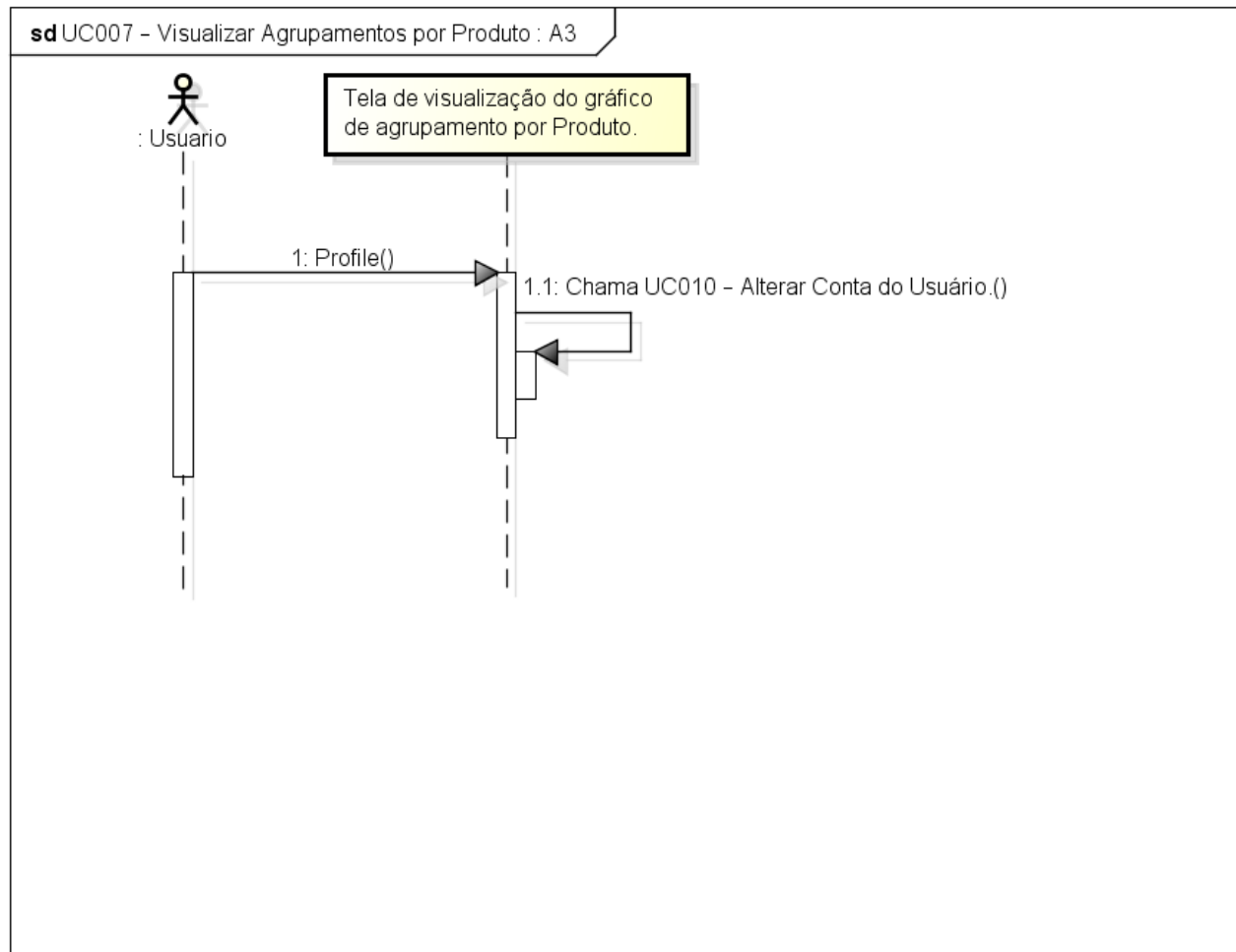


FIGURA 43 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A3
 FONTE: Os autores (2013)

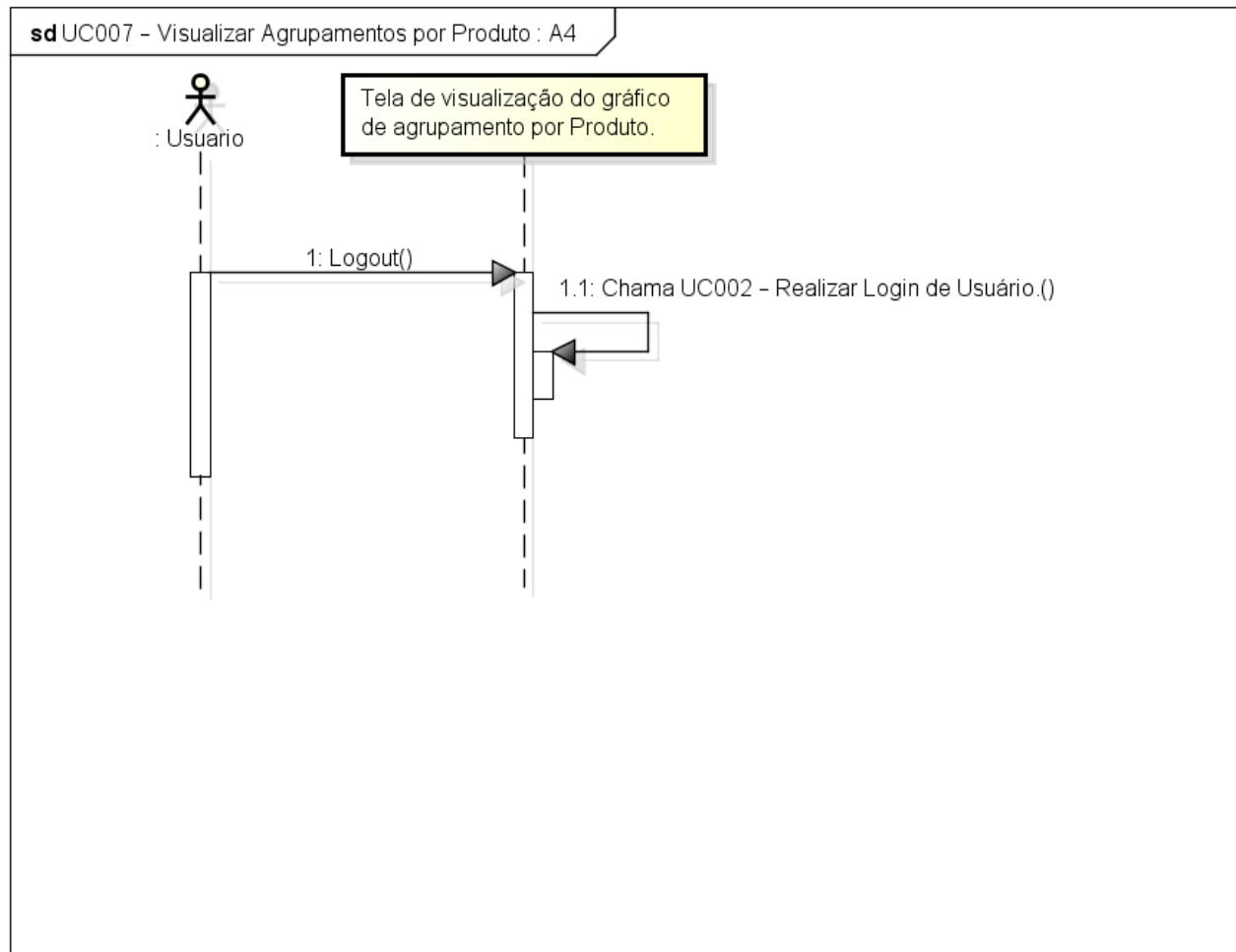


FIGURA 44 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A4
FONTE: Os autores (2013)

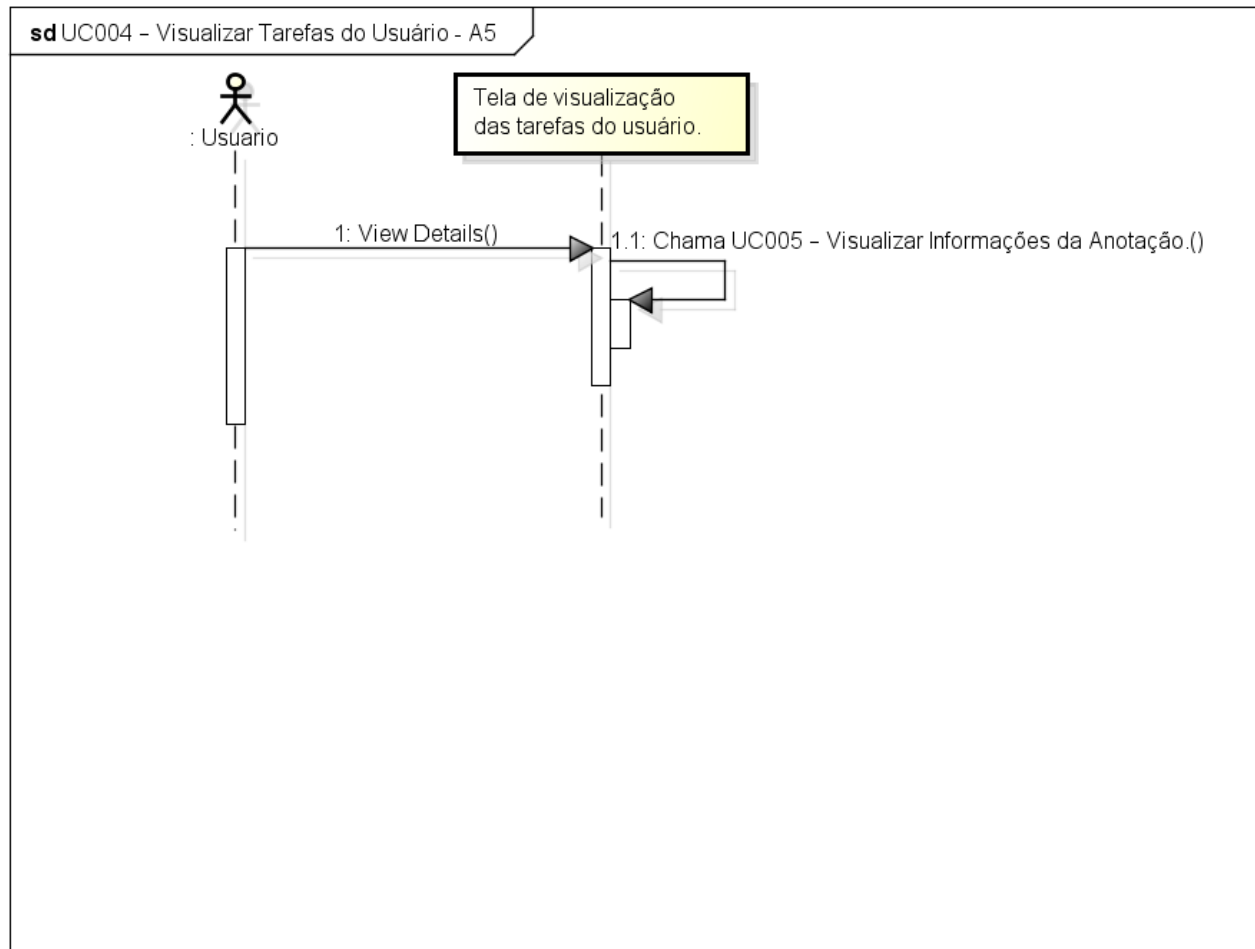


FIGURA 45 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR PRODUTO - FLUXO A5
FONTE: Os autores (2013)

8.8 UC008 - Visualizar Agrupamentos por Organismo

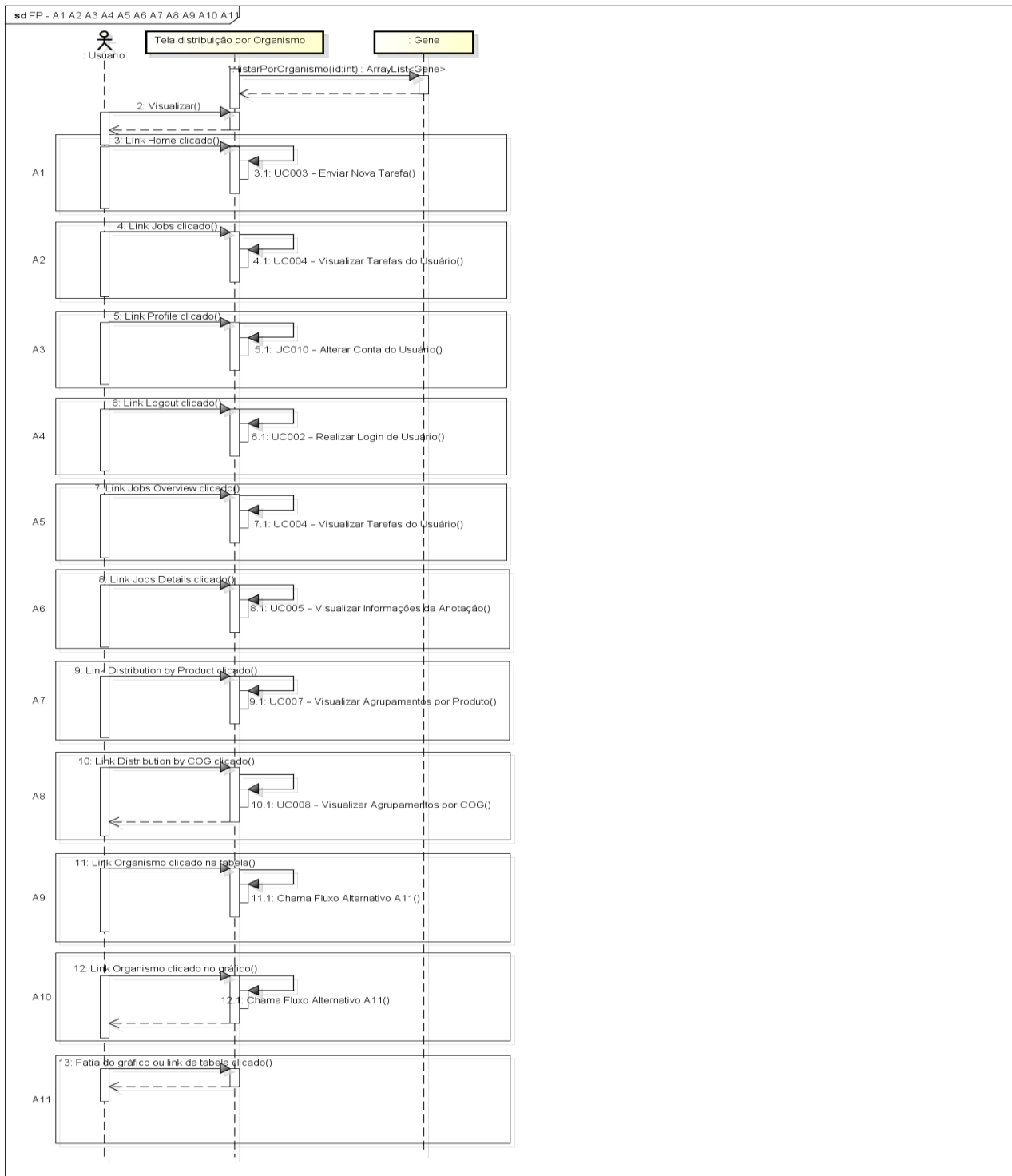


FIGURA 46 - DIAGRAMA DE SEQUENCIA - VISUALIZAR AGRUPAMENTOS POR ORGANISMO - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10 e A11

FONTE: Os autores (2013)

8.9 UC009 - Visualizar Informação do gene

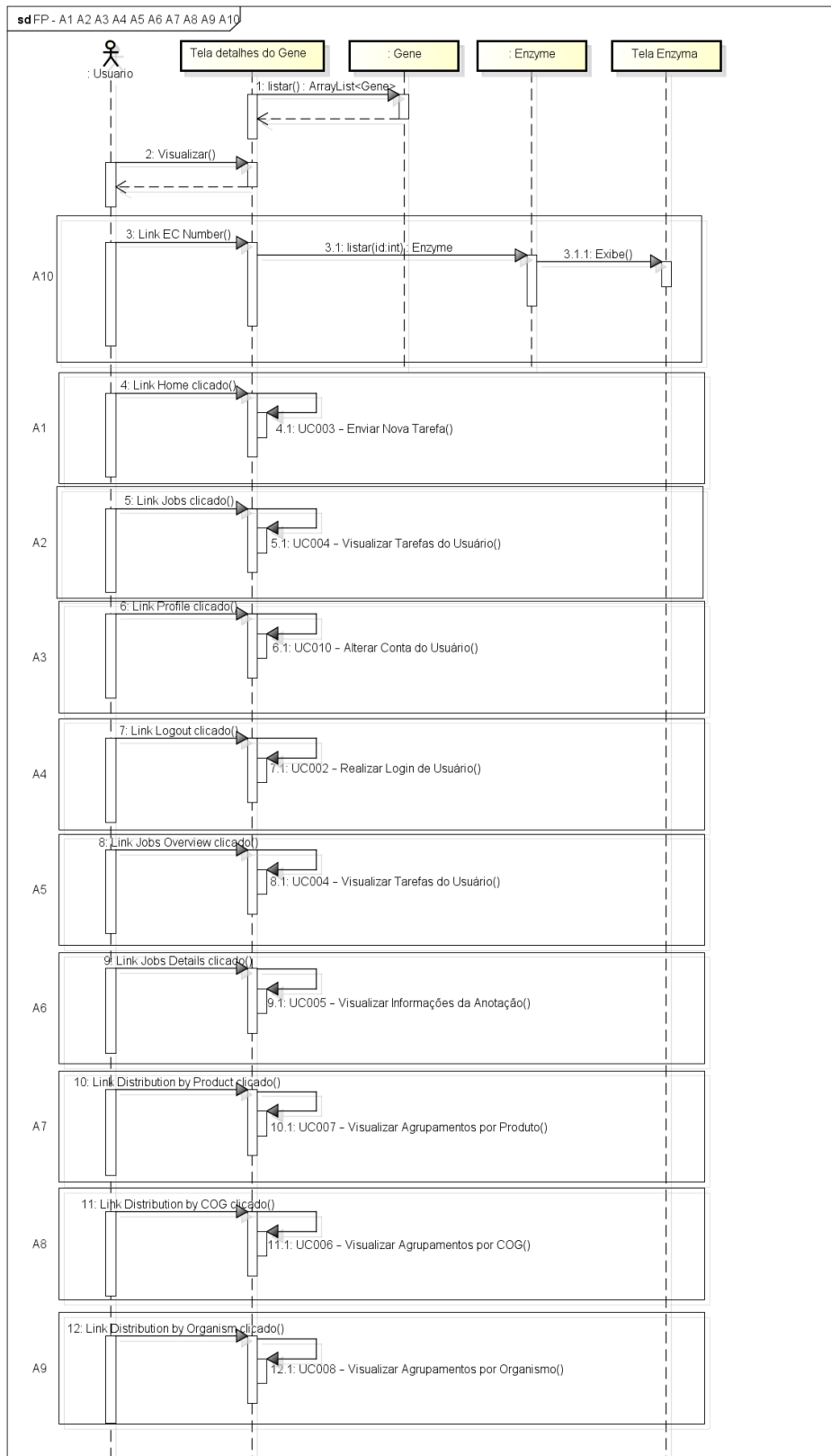


FIGURA 47 - DIAGRAMA DE SEQUENCIA - VISUALIZAR INFORMAÇÃO DO GENE - FLUXO PRINCIPAL, A1, A2, A3, A4, A5, A6, A7, A8, A9 e A10

FONTE: Os autores (2013)

8.10 UC010 - Alterar Conta do Usuário

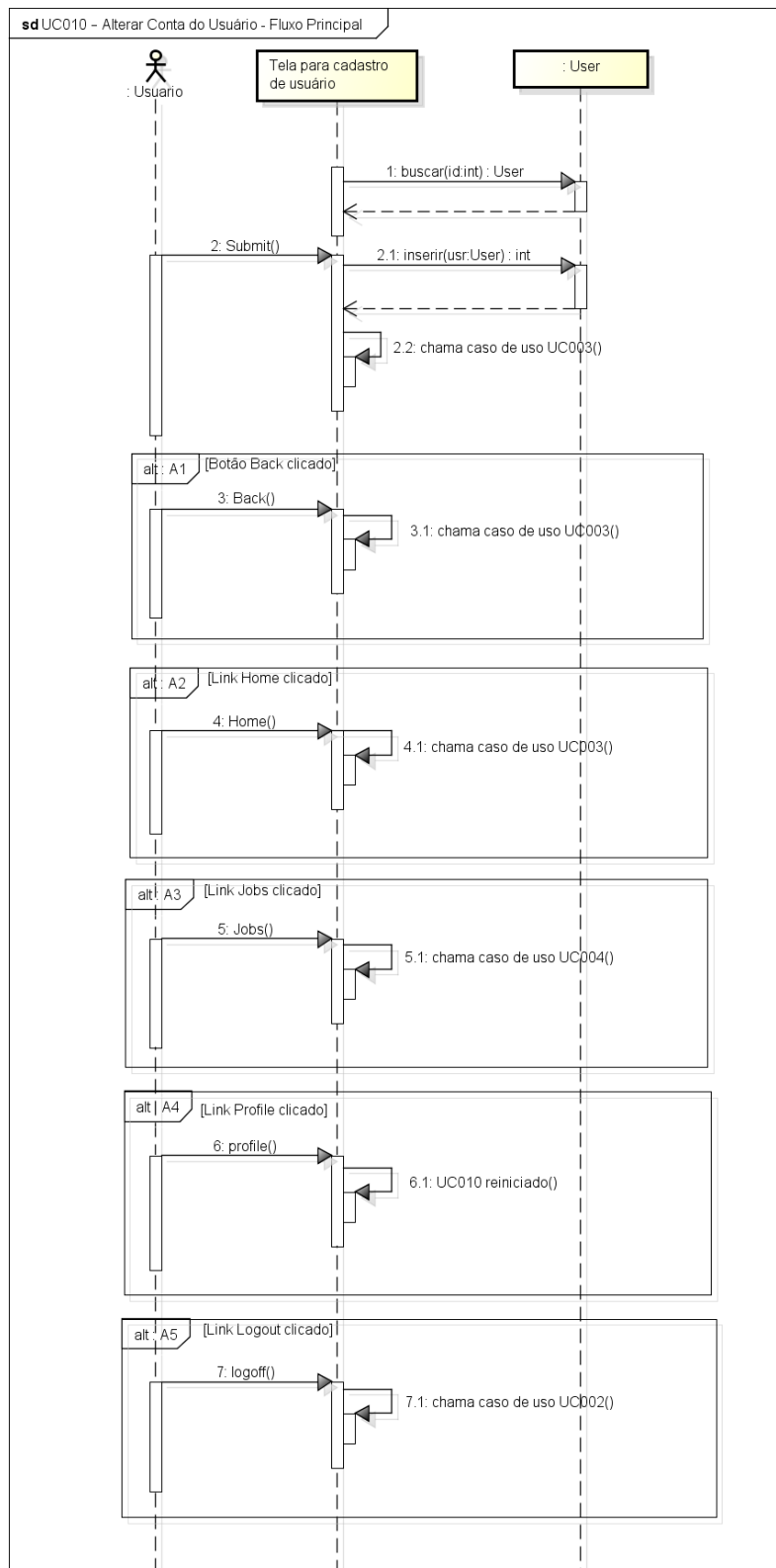


FIGURA 48 - DIAGRAMA DE SEQUENCIA - ALTERAR CONTA DO USUÁRIO - FLUXO PRINCIPAL, A1, A2, A3, A4 e A5

FONTE: Os autores (2013)

9 PGP – PLANO DE GERENCIAMENTO DE PROJETO

9.1 INTRODUÇÃO

9.1.1 ESCOPO E PROPÓSITO DO DOCUMENTO

Este documento reúne os procedimentos necessários para o desenvolvimento de um sistema web integrado ao software de anotação SILA para fornecimento de seus serviços a visualização das informações. Aqui serão descritos a forma de gerenciamento e organização necessários para a execução do projeto.

9.1.2 OBJETIVOS DO PROJETO

9.1.2.1 OBJETIVOS

O projeto do sistema tem como objetivo oferecer um serviço capaz de auxiliar alguns processos realizados em laboratórios de bioinformática, na área de anotação de genomas. Terá como objetivo principal integrar as funcionalidades da ferramenta de anotação automática SILA disponibilizando a sua operação as informações geradas pela internet.

9.1.2.2 FUNÇÕES PRINCIPAIS

O sistema gerência o fornecimento da entrada de arquivos texto padrão (Fasta) para o software SILA e processa os resultados gerados no formato GenBank exibindo as informações em ambiente web:

- Fornecer uma interface navegável para execução das tarefas;
- Gerenciar contas de usuários;
- Gerenciar tarefas de usuários;
- Fornecer diferentes opções de visualização dos dados de anotação;

9.1.3 RESTRIÇÕES TÉCNICAS E ADMINISTRATIVAS

O projeto possui tempo de execução curto, um período de 5 meses. Por este motivo não é possível aplicar as funcionalidades necessárias para o problema de anotação de genomas, tais como a possibilidade de alterações manuais na anotação. Esta limitação pode ser contornada com uma continuidade no projeto. Há uma limitação técnica no conhecimento da área de biologia molecular e bioinformática dos participantes, que terão que dividir o tempo de projeto e desenvolvimento com tempo de estudo dos conceitos e problemas da área.

9.1.4 ESTIMATIVAS DO PROJETO

9.1.4.1 DADOS HISTÓRICOS USADOS NAS ESTIMATIVAS

As estimativas do software foram baseadas em experiências profissionais dos envolvidos na área de desenvolvimento de software e planejamento. Para avaliar a complexidade de cada item, utilizamos valores padrão e nos remetemos aos trabalhos que foram previamente realizados.

A margem de erro desta estimativa é média/alta, pois a maioria dos envolvidos não possuem experiência no desenvolvimento do softwares para bioinformática.

9.1.4.2 TÉCNICAS DE ESTIMATIVAS

A técnica utilizada para estimar o tempo para realização do software foi a análise de pontos por caso de uso. Esta técnica baseia as estimativas nos casos de uso e atores utilizados na modelagem do projeto (imagem UC).

Os elementos do sistema são classificados e atribuídos a pontuações conforme sua complexidade. Estes valores geram pontos de caso de uso não ajustados. Também são contabilizados fatores de complexidade técnica (características que podem atribuir dificuldades no desenvolvimento do software) e ambiental (nível de experiência dos

profissionais envolvidos). Todos estes valores calculados resultam em pontos de caso de uso ajustados. Utilizando a margem de 28 homens-hora, é estimada a duração do projeto.

9.1.4.3 ESTIMATIVAS

Total de Pesos não Ajustados de Atores

Atores	Nível de complexidade
Usuário	3
Total	
Simple (Total: 0 – Peso: 1)	0
Médio (Total: 0 – Peso: 2)	0
Complexo (Total: 1 – Peso: 3)	3
TPNAA	3

Total de Pesos não Ajustados dos Casos de Uso

Caso de uso	Nível de complexidade
Gerenciar Cadastro de Usuário	1
Realizar Login de Usuário	1
Enviar Nova Tarefa	2
Visualizar Tarefas do Usuário	1
Visualizar mapa do genoma	3
Visualizar Agrupamentos por Produto	3
Visualizar Agrupamentos por COG	3
Visualizar Agrupamentos por Organismo	3
Total	
Simple (Total: 3 – Peso: 1)	3
Médio (Total: 1 – Peso: 2)	2
Complexo (Total: 4 – Peso: 3)	12
TPNAUC	17

Fator de complexidade técnica

Descrição	Peso	Valor	Total
Sistemas Distribuídos	2	1	2
Desempenho da Aplicação	1	4	4
Eficiência do usuário final	1	5	5
Processamento interno complexo	1	4	4
Reusabilidade do código	1	3	3
Facilidade de Instalação	0,5	4	2
Usabilidade	0,5	5	2,5
Portabilidade	2	3	6
Manutenibilidade	1	5	5
Concorrência	1	3	3
Características especiais de segurança	1	5	5
Acesso direto para terceiros	1	5	5
Facilidades especiais de treinamento	1	2	2
Soma	48,5		
FCT = 0,6 + (0,01 * Soma) = 1,085			

Fator de complexidade ambiental

Descrição	Peso	Valor	Total
F1 Familiaridade com o processo de desenvolvimento de software	1,5	2	3
F2 Experiência na aplicação	0,5	2	1
F3 Experiência com OO, na linguagem e na técnica de desenvolvimento	1	2	2
F4 Capacidade do líder de análise	0,5	2	1
F5 Motivação	1	3	3
F6 Requisitos estáveis	2	1	2
F7 Trabalhadores com dedicação parcial	-1	4	-4
F8 Dificuldade na linguagem de programação	-1	4	-4
Soma		4	
FCA = $1,4 + (-0,03 * Soma) = 1,28$			

Pontos por caso de uso não ajustados:

- $PCUNA = TPNA A + TPNAUC = 20$

Pontos por caso de uso ajustados:

- $PCUA = PCUNA * FCT * FCA = 27,776$

Estimativa de horas:

- $28hh * 27,776 = 777,728horas = 25,9242 \text{ dias}$

9.1.5 RISCOS DO PROJETO

9.1.5.1 ANÁLISE DOS RISCOS

9.1.5.2 IDENTIFICAÇÃO

1. Identificação de bibliotecas: o projeto irá necessitar de bibliotecas para o desenvolvimento de algumas funcionalidades. São bibliotecas que implementem visualizações gráficas e bibliotecas que manipulem arquivos da área.

2. Pessoal: a equipe possui 4 integrantes e é dependente das suas funções para o projeto inteiro, tendo um cronograma justo para a realização das tarefas.

3. Linguagem de programação: o conhecimento avançado em alguns assuntos da linguagem utilizada (Java) será necessário, sendo necessário estudo prévio. É possível também que a linguagem não consiga suprir uma necessidade da maneira como esperado.

4. Tempo: o tempo destinado ao planejamento e desenvolvimento do projeto não pode ser ultrapassado.

5. Conhecimento limitado na área: por ser um software de bioinformática, é necessário um conhecimento de alguns conceitos previamente para execução do projeto.

6. Desempenho: o sistema irá demandar intenso uso do hardware e deve ser planejado para conseguir atingir esta meta. Isto pode gerar atrasos na análise de dos resultados a serem visualizados.

9.1.5.3 AVALIAÇÃO DOS RISCOS

1. Identificação de bibliotecas:

Data limite: 28/12/2012

Consequência: Atraso no início do desenvolvimento de algumas funcionalidades.

Ação: Encontrar funcionalidades alternativas ou programar a biblioteca.

Monitoramento: ao inicio do desenvolvimento e na implementação das bibliotecas

Probabilidade: Muito baixa

Impacto: Muito alto

2. Pessoal:

Data limite: Todo o projeto

Consequência: Possível atraso em algum estágio

Ação: Alterar cronograma conforme dificuldades inesperadas acontecem, possibilitando que os estágios sejam finalizados.

Monitoramento: Constante, semanal

Probabilidade: Moderada

Impacto: Baixo / Moderado

3. Linguagem de programação:

Data limite: Todo o projeto

Consequência: Possível atraso em algum estágio, impossibilidade de realizar funcionalidades

Ação: rever a funcionalidade de acordos com as limitações da linguagem, buscar métodos alternativos. Manter estudo na área da linguagem de programação utilizada para evitar outros riscos

Monitoramento: um estágio de antecedência do desenvolvimento

Probabilidade: Moderada

Impacto: Alto

4. Tempo:

Data limite: 20/03/2013

Consequência: Atraso na entrega

Ação: Simplificar estágio e/ou funcionalidade para ajustar tempo de desenvolvimento. Dedicar mais tempo ao projeto

Monitoramento: Constante verificação do cronograma

Probabilidade: Moderada

Impacto: Alto

5. Conhecimento limitado na área:

Data limite: Todo o projeto

Consequência: Não compreensão dos problemas, dificuldades no desenvolvimento

Ação: Realizar estudo prévio, identificar todo o conteúdo que envolva o projeto, reunião com orientador para dúvidas

Monitoramento: Constante

Probabilidade: Alta

Impacto: Baixo

6. Desempenho:

Data limite: Ao final de cada estágio

Consequência: remoção de funcionalidade, software com tarefas lentas, atraso na entrega

Ação: Identificar os pontos críticos e revisar bibliografia para melhorar o desempenho. Utilizar outros métodos

Monitoramento: Durante as tarefas de desenvolvimento do projeto

Probabilidade: Alta

Impacto: Alto no desempenho, moderado no atraso

9.1.5.4 ADMINISTRAÇÃO DOS RISCOS

9.1.5.5 OPÇÕES PARA EVITAR OS RISCOS

1. Identificação de bibliotecas: pesquisa com antecedência de bibliotecas que realizam as funções desejadas. Encontrar várias opções, tendo redundância e opção de escolha. Verificar bibliotecas comuns na área, as mais utilizadas e recentes, com documentação atualizada.

2. Pessoal: planejar cronograma com mais tempo para tarefas críticas.

3. Linguagem de programação: revisar bibliografia da linguagem nos assuntos que serão mais utilizados. Analisar programas semelhantes e suas técnicas.

4. Tempo: ser cauteloso na distribuição de tempo. Integrantes devem estar cientes que podem precisar de mais tempo do que o normal para finalização do projeto, abrindo mão de outras atividades.

5. Conhecimento limitado na área: realizar estudo e ter visão geral do problema a ser desenvolvido. Procurar assistência com orientador do projeto e profissionais da área para sanar dúvidas. Dividir estudos por estágios, conforme necessidades aparecem.

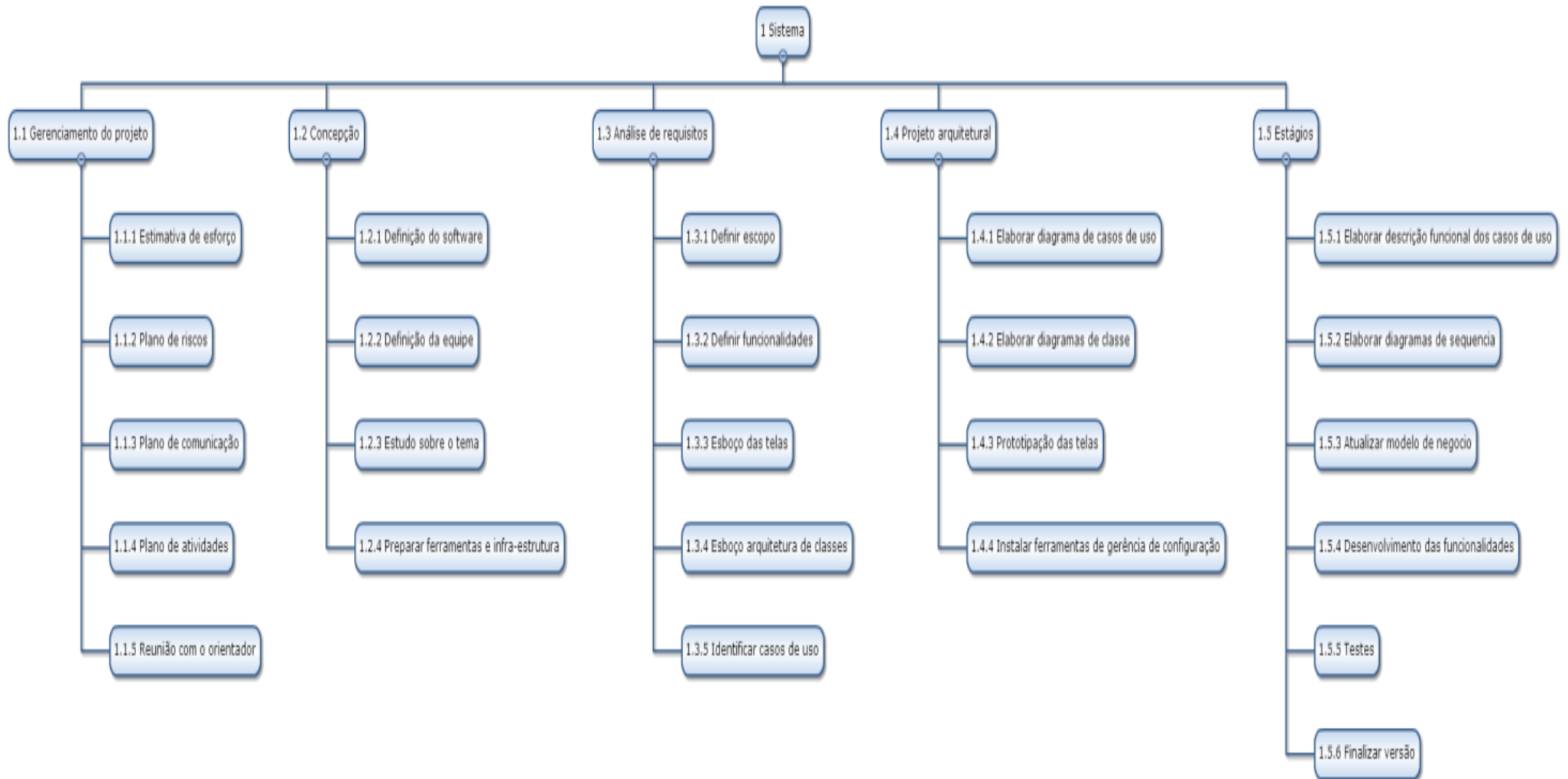
6. Desempenho: verificar técnicas de softwares semelhantes. Procurar métodos alternativos para solução de problemas. Rever funcionalidade.

9.1.5.6 PROCEDIMENTOS DE MONITORAÇÃO DOS RISCOS

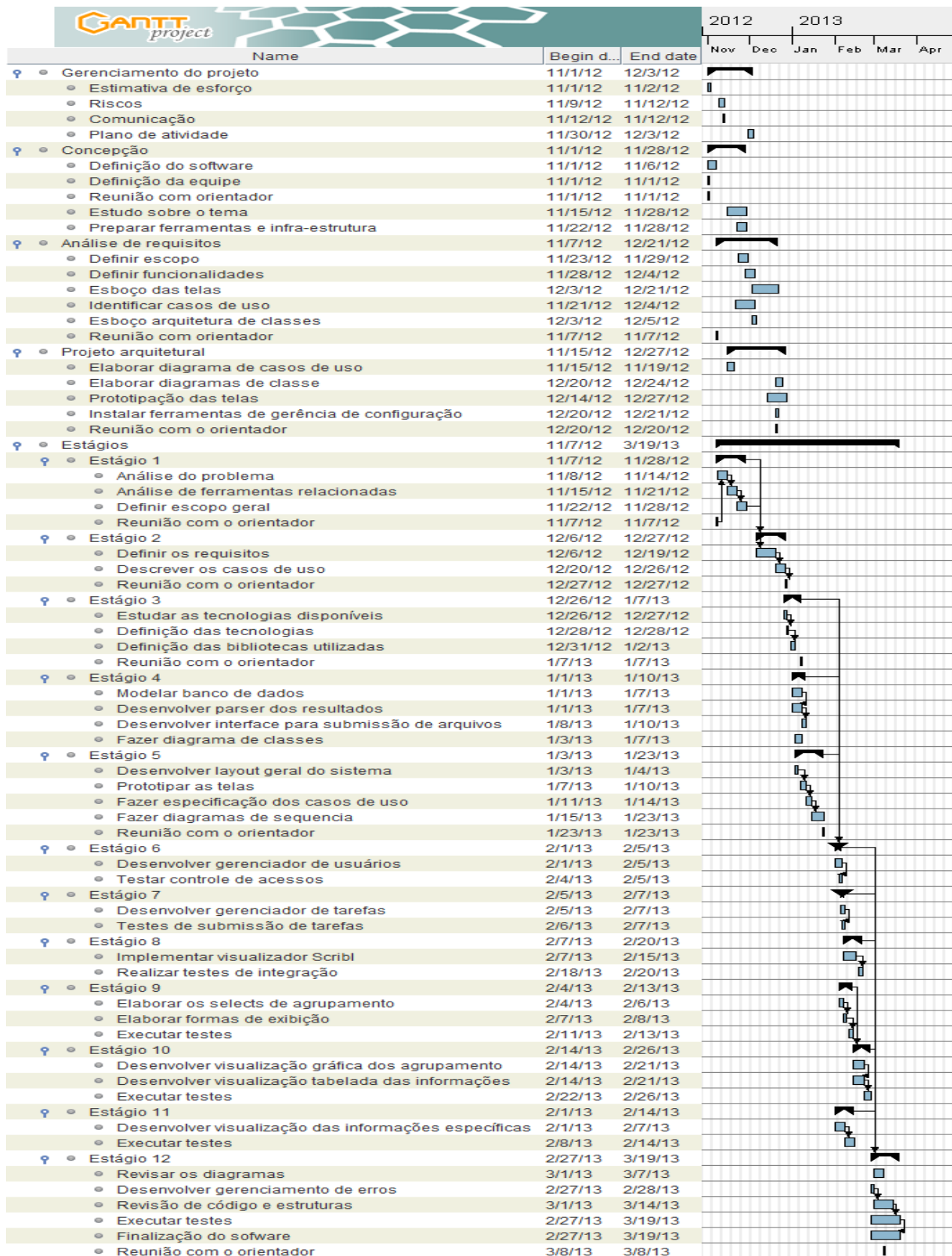
Serão realizadas reuniões semanais para avaliar o andamento e avaliar possíveis riscos no decorrer do projeto. Desta maneira será mais fácil a rápida identificação do problema, encontrando uma possível solução, que não resulte em atraso nem problemas no software.

9.2 CRONOGRAMA

9.2.1 WORK BREAKDOWN STRUCTURE



9.2.2 GRÁFICO DE GANTT



9.3 RECURSOS DO PROJETO

9.3.1 PESSOAL

A equipe será formada por 4 membros: Fábio Roberto Stella Santos, Fábio Henrique Jugler, Jackson Augusto Probst e Ricardo Assunção Vialle, tendo o Professor Roberto Tadeu Raittz como professor orientador.

9.3.2 HARDWARE E SOFTWARE

Para realização do projeto, desenvolvimento, testes e documentação serão utilizados 4 notebooks:

- Processador Intel Core 2 Duo (2.2GHz), Memória RAM 4GB, HD 500GB
- Processador Intel Core i7 (1.73GHz), Memória RAM 12GB, HD 756GB
- Processador Intel Core i3 (2.3GHz), Memória RAM 3GB, HD 320GB
- Processador Intel Core i5 (2.4GHz), Memória RAM 4GB, HD 500GB

O programa será desenvolvido na plataforma JAVA, JDK versão 6, utilizando os seguintes softwares (para desenvolvimento, planejamento e documentação):

- Sistema operacional: Windows Seven
- IDE de desenvolvimento e testes: Eclipse Java EE IDE for Web Developers
- Documentação: Microsoft Office Professional 2007
- Diagramas: Astah Community 6.5.1, GanttProject 2.6.1, WBSTool.com.

* Obs: Os softwares que necessitam de licença paga foram utilizados em suas respectivas versões de avaliação, com licença para estudantes.

9.4 ORGANIZAÇÃO DO PESSOAL

9.4.1 ESTRUTURA DA EQUIPE

Todos os integrantes se dividem no trabalho de análise e desenvolvimento do sistema além de contribuírem na parte estrutural e de testes. A divisão do trabalho se estende na parte da gerencia e acompanhamento do projeto e construção de diagramas.

Roberto como professor orientador, irá auxiliar no andamento do projeto e possíveis dúvidas.

9.5 PLANO DE COMUNICAÇÃO

Grupos interessados	Foco	O que este grupo precisa saber	Método	Quando	Observação
Equipe	Compartilhar Conhecimento Dificuldades e aprendizados	Quais as dúvidas? Onde pesquisar? O que pesquisar?	Troca de conteúdo estudado	Diariamente	Quem tiver algo a compartilhar (dúvidas, descobertas, realizações) compartilha via e-mail ou outro meio disponível
Equipe	Feedback das atividades	O que desenvolveu	Reunião de 2 horas	Semanalmente	Envolvidos enviam e-mails confirmando presença.
Equipe + Professor orientador	Tirar dúvidas Verificar pendências Verificar funcionalidades	Dificuldades e problemas encontrados Soluções encontradas	Reunião de 2 horas	Quinzenalmente	Responsável agenda horário com professor orientador
Equipe + Usuário especializado	Testar a aplicação com dados reais de laboratório	Usar o software Obter os arquivos para teste	Encontro em laboratório da UFPR para uso do software	Dependendo da disponibilidade e do usuário, reunião ocorre na fase final do projeto	Responsável agenda horário com usuário especializado

9.6 MECANISMOS DE RASTREAMENTO E CONTROLE

Os controles do projeto são de responsabilidade de todos os envolvidos. Os aspectos a serem controlados estão indicados abaixo.

9.6.1 VERIFICAÇÃO E CONTROLE DO ESCOPO

Para a verificação do escopo serão utilizados check lists em reuniões periódicas a fim de manter o projeto nas conformidades do escopo. Possíveis alterações de projeto

que possam impactar no escopo deverão ser atualizadas no check-list para que a documentação seja refatorada.

9.6.2 CRONOGRAMA

O controle do cronograma será executado com o auxílio da lista de atividades e o do gráfico de Gantt. Com isto será possível o acompanhamento das atividades bem como seus possíveis atrasos, permitindo o remanejamento rápido de recursos para evitar atrasos excessivos. Como o projeto se trata de um trabalho acadêmico, existe um limite nos recursos humanos empregados, não existindo a possibilidade de agregação de novos recursos quando necessário, isso torna o cronograma um risco.

9.6.3 QUALIDADE

A verificação da qualidade do projeto será obtida por testes periódicos do software pelo professor orientador e por uma equipe especialista por ele indicada. Assim teremos retorno da qualidade e andamento do desenvolvimento.

9.6.4 GERENCIAMENTO DA EQUIPE DO PROJETO

Será realizado por cada membro da equipe uma vez que não existe uma hierarquia dentre os participantes do projeto. Os membros deverão observar uns aos outros para garantir o comprometimento e o cumprimento das exigências do projeto. A saída repentina de membros deverá ser sanada com a agregação de membros substitutos o mais rápido possível para que não aconteçam grandes impactos no cronograma.

9.6.5 RELATÓRIO DE DESEMPENHO

O desempenho do andamento do projeto será analisado em reuniões periódicas. Assim será possível analisar as metas estabelecidas e atingidas para cada um dos indicadores do projeto.

9.6.6 GERENCIAMENTO DAS PARTES INTERESSADAS

Reuniões deverão acontecer com o orientador do projeto para que se garanta o sucesso na entrega do mesmo. As opiniões e solicitações do coordenador serão de grande relevância uma vez que o projeto será avaliado por uma banca de professores ao final do prazo estimado.

9.6.7 RISCOS

O propósito da monitoração dos riscos é determinar se as respostas aos riscos estão sendo implementadas como planejadas e tão efetivas quanto esperadas. Também visa se devem ser desenvolvidas novas respostas, bem como se as premissas do projeto ainda são válidas, e se a exposição ao risco mudou frente ao seu estado anterior. O acompanhamento deve observar se um gatilho de risco ocorreu e se as políticas e procedimentos adequados estão sendo seguidos. O relatório de riscos será avaliado periodicamente nas reuniões para este fim e as ações serão conduzidas de acordo com as conclusões e decisões das reuniões de acompanhamento.

9.6.8 MUDANÇA DE ESCOPO

Qualquer alteração que se fizer necessária e que implique na alteração de um dos documentos do projeto, será considerada como mudança de escopo. A solicitação deverá ser analisada por todos os membros da equipe para medir o impacto da mudança no projeto e definição da recomendação a ser seguida. A implementação de uma mudança somente acontecerá, se recomendada pela equipe e aprovada pelo orientador do projeto.